

A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes II: the importance of mRNA secondary structure in assessing the functionality of 3' UTR variants

Jian-Min Chen · Claude Férec · David N. Cooper

Received: 3 April 2006 / Accepted: 29 May 2006 / Published online: 29 June 2006
© Springer-Verlag 2006

Abstract In an attempt both to catalogue 3' regulatory region (3' RR)-mediated disease and to improve our understanding of the structure and function of the 3' RR, we have performed a systematic analysis of disease-associated variants in the 3' RRs of human protein-coding genes. We have previously analysed the variants that have occurred in two specific domains/motifs of the 3' untranslated region (3' UTR) as well as in the 3' flanking region. Here we have focused upon 83 known variants within the upstream sequence (USS; between the translational termination codon and the upstream core polyadenylation signal sequence) of the 3' UTR. To place these variants in their proper context, we first performed a comprehensive survey of known

cis-regulatory elements within the USS and the mechanisms by which they effect post-transcriptional gene regulation. Although this survey supports the view that RNA regulatory elements function within the context of specific secondary structures, there are no general rules governing how secondary structure might exert its influence. We have therefore addressed this question by systematically evaluating both functional and non-functional (based upon *in vitro* reporter gene and/or electrophoretic mobility shift assay data) USS variant-containing sequences against known *cis*-regulatory motifs within the context of predicted RNA secondary structures. This has allowed us not only to establish a reliable and objective means to perform secondary structure prediction but also to identify consistent patterns of secondary structural change that could potentiate the discrimination of functional USS variants from their non-functional counterparts. The resulting rules were then used to infer potential functionality in the case of some of the remaining functionally uncharacterised USS variants, from their predicted secondary structures. This not only led us to identify further patterns of secondary structural change but also several potential novel *cis*-regulatory motifs within the 3' UTRs studied.

Electronic Supplementary Material Supplementary material is available to authorised users in the online version of this article at <http://dx.doi.org/10.1007/s00439-006-0218-x>.

J.-M. Chen · C. Férec
INSERM, U613, 29220 Brest, France

J.-M. Chen (✉) · C. Férec
Etablissement Français du Sang – Bretagne,
46 rue Félix Le Dantec, 29220 Brest, France
e-mail: Jian-Min.Chen@univ-brest.fr

J.-M. Chen · C. Férec
Faculté de Médecine de Brest et des Sciences de la Santé,
Université de Bretagne Occidentale, 29238 Brest, France

C. Férec
Laboratoire de Génétique Moléculaire et
d'Histocompatibilité, CHRU Brest, Hôpital Morvan,
29220 Brest, France

D. N. Cooper
Institute of Medical Genetics, Cardiff University,
Heath Park, Cardiff CF14 4XN, UK

Introduction

In an attempt both to catalogue 3' regulatory region (3' RR)-mediated disease and to improve our understanding of the structure and function of the 3' RR, we have performed a systematic analysis of disease-associated variants in the 3' RRs of human protein-coding genes. In a previous article, we discussed the general

principles pertinent to this meta-analytical study and focused on the variants that are known to have occurred in two specific domains/motifs [i.e. the upstream core polyadenylation signal sequence (UCPAS) and the left arm of the ‘spacer’ sequence (LAS) between the UCPAS and the pre-mRNA cleavage site (CS)] of the 3′ untranslated region (3′ UTR) as well as in the 3′ flanking region (Chen et al. 2006). In the present article, we shall focus on the 83 known variants that are known to have occurred within the upstream sequence (USS) between the translational termination codon and the UCPAS of the 3′ UTR (refer to Fig. 1 in Chen et al. 2006 for definition of terms).

Classification of the USS variants

The 81 previously collated USS variants (Chen et al. 2006) were essentially derived from the Human Gene Mutation Database (<http://www.hgmd.org>; Stenson et al. 2003) with the inclusion of several entries found by cross-reference search (note that no public polymorphism databases were used for data collection). In addition, two recently published and functionally characterised USS variants in the *BMP2* (Fritz et al. 2006) and *SEPN1* (Allamand et al. 2006) genes were also included for analysis (Table 1; see Supplementary Tables S1–S5 for sequence details). Although the general principles involved in analysing the 3′ RR variants have been previously described (Chen et al. 2006), it is nevertheless important to explain how the USS variants were sub-divided into the five groups listed in Table 1. The first two groups were assigned essentially on the basis of the results of either in vitro reporter gene and/or electrophoretic mobility shift assay (EMSA) data: a variant was termed ‘functional’ when it exerted a marked effect on gene expression (at either the mRNA or protein level) in a reporter gene assay and/or affected *trans*-acting factor binding in the EMSA assay; otherwise, variants were termed ‘non-functional’. The third group contains those variants which have not been characterised in vitro but which have supporting in vivo gene expression data whereas the fourth group constitutes those variants that lack any in vitro or in vivo supporting data. Finally, four variants were treated as isolated examples owing to their complex nature.

Analytical strategy

It has long been appreciated that many RNA regulatory motifs rely for their function on a combination of

primary and secondary structure (e.g. Mignone et al. 2002). However, to date, no general rules have been formulated that govern how RNA secondary structure might operate to mediate the functionality of the component regulatory elements. Here we have set out to address this basic question by systematically evaluating both functional and non-functional USS variant-containing primary sequences against the well-defined *cis*-regulatory motifs and within the context of the predicted RNA secondary structures. The rules derived from this study were then used to infer potential functionality, in the case of some of the remaining functionally uncharacterised USS variants, from their respective secondary structures. First, however, a survey of known *cis*-regulatory elements within the USS is warranted.

Overview of well-defined regulatory elements in the USS

As we shall see below, a diverse range of *cis*-acting elements has been identified in the USS. Note that (a) in the interest of brevity, we have focused primarily on human genes. Other systems and organisms have been discussed only where data from humans were limited or absent; (b) usually only well-characterised elements (i.e. those mapped within shorter RNA fragments and experimentally validated) were collated; (c) secondary structural features were discussed whenever such information was available in the original publications; and (d) although the examples given are by no means comprehensive (despite an extensive literature search), taken together they represent all the major stages of post-transcriptional gene regulation.

Auxiliary regulatory elements involved in mRNA 3′ end formation

In addition to the three core elements required for polyadenylation [i.e. UCPAS, DCPAS (downstream core polyadenylation signal), and CS; refer to Fig. 1 in Chen et al. 2006], other sequence elements have also been identified that modulate the efficiency of mRNA 3′ end formation. One class of such auxiliary regulatory elements is located upstream of the UCPAS hexamer and are thus termed upstream sequence elements (USEs; reviewed in Zhao et al. 1999). Human USEs are generally U-rich, but a consensus sequence has not yet been established (Table 2). U-rich USEs in human genes bind to Fip1, a subunit of cleavage and polyadenylation specificity factor (CPSF), thereby contributing to CPSF-mediated stimulation of poly(A) polymerase activity (Kaufmann et al. 2004).

Table 1 Summary of USS variants analysed in this study

Disease(s)/trait(s) under investigation ^a	Gene	Variant ^b	Main reference(s)	Predicted secondary structure change (and primary sequence or motif involved where appropriate) ^c
Group 1: Non-functional variants (n = 4)				
– ^d	<i>CXCL12</i>	G801A polymorphism	Winkler et al. (1998), Arya et al. (1999), Kimura et al. (2003)	Pattern 0
–	<i>APOA5</i>	T>C SNP ^e	Pennacchio et al. (2001), Talmud et al. (2005)	Pattern 0
–	<i>NPPC</i>	G>A polymorphism	Ono et al. (2002)	Pattern 0
–	<i>SEPI5</i>	811C>T polymorphism	Kumaraswamy et al. (2000), Hu et al. (2001)	Pattern 0
Group 2: Functional variants (n = 17)				
β -thalassaemia intermedia	<i>HBB</i>	C>G mutation	Jankovic et al. (1991), Maragoudaki et al. (1998), Sgourou et al. (2002)	Pattern I involving CU-rich sequence
β -thalassaemia	<i>HBB</i>	13 bp deletion mutation	Basak et al. (1993), Bilenoglu et al. (2002)	–
mRNA stability/expression	<i>BMP2</i>	A>C SNP	Fritz et al. (2006)	Pattern II involving AU-rich sequence
Rheumatoid arthritis	<i>NR3C1</i>	A>G polymorphism	Derijk et al. (2001), Schaaf and Cidowski (2002)	Pattern II involving AU-rich sequence
Arrhythmogenic right ventricular dysplasia	<i>TGFB3</i>	1723 C>T mutation	Beffagna et al. (2005)	Pattern II involving AU-rich sequence
Tumour onset and progression	<i>CEACAM1</i>	Single T deletion polymorphism	Ruggiero et al. (2003)	Pattern I involving U-rich sequence
Insulin resistance and type 2 diabetes	<i>PPPIR3A</i>	5 bp deletion polymorphism	Xia et al. (1998, 1999)	–; involving AUUUA motif
Cancer susceptibility; sensitivity to chemotherapy and preoperative chemoprevention	<i>TYMS</i>	6 bp deletion polymorphism	Ulrich et al. (2000), Graziano et al. (2004), Mandola et al. (2004), Liao et al. (2006), Zhang et al. (2005), Lu et al. (2006)	–; involving AU-rich sequence
Blood pressure	<i>NPRI</i>	Single C deletion polymorphism	Knowles et al. (2003), Pitzalis et al. (2003)	Pattern II involving C-rich sequence
Alzheimer disease	<i>TFCP2</i>	G/A polymorphism	Lambert et al. (2000)	Pattern I involving CU-rich sequence
mRNA instability	<i>CYP2A6</i>	Two SNPs	Wang et al. (2006)	–; involving AG-rich sequences
F7 deficiency	<i>F7</i>	AA insertion polymorphism	Peyvandi et al. (2005)	Pattern II involving AG-rich sequence
Insulin resistance	<i>PTPN1</i>	G insertion polymorphism	Di Paola et al. (2002)	Pattern I involving AG-rich sequence
Gene expression	<i>NPRI</i>	AGAA deletion polymorphism	Knowles et al. (2003)	Pattern I involving AG-rich sequence
Several types of cancer	<i>SEPI5</i>	G/A1125 polymorphism	Kumaraswamy et al. (2000), Hu et al. (2001)	Pattern I
Rigid spine muscular dystrophy	<i>SEPN1</i>	T>C mutation	Allamand et al. (2006)	Pattern II involving a hairpin loop comprising of CCA(U/C)GAU
Tourette syndrome	<i>SLITRK1</i>	G>A mutation	Abelson et al. (2005)	Pattern I
Group 3: Variants with supporting in vivo data (n = 12)				
Biotinidase deficiency	<i>BTD</i>	G>A mutation	Muhl et al. (2001)	Pattern I involving AG-rich sequence

Table 1 continued

Disease(s)/trait(s) under investigation ^a	Gene	Variant ^b	Main reference(s)	Predicted secondary structure change (and primary sequence or motif involved where appropriate) ^c
Cholesteryl ester transfer protein activity	<i>CETP</i>	G>A polymorphism	Tamminen et al. (1996)	Pattern I
Systematic lupus erythematosus	<i>CRP</i>	CRP3 and CRP4 polymorphisms	Russell et al. (2004)	CRP3: pattern II involving a loop comprising AAA(C/U)GG; CRP4: pattern 0
Glucose intolerance and type 2 diabetes	<i>ENPP1</i>	A>G polymorphism	Meyre et al. (2005)	Pattern I involving CU-rich sequence
Type 2 diabetes mellitus	<i>GFP72</i>	SNP10 (G/A)	Zhang et al. (2004)	Pattern III
Hexosaminidase activity	<i>HEXB</i>	TG deletion polymorphism	Kleiman et al. (1998)	Pattern II involving a hairpin loop (CUGUUUU>CUUUU)
Histamine <i>N</i> -methyltransferase activity	<i>HNNMT</i>	A939G polymorphism	Preuss et al. (1998)	Pattern I involving AU-rich sequence
Body weight	<i>IGF2</i>	G>A polymorphism	O'Dell et al. (1997), Gaunt et al. (2001), Roth et al. (2002), Schrager et al. (2004), Gomes et al. (2005)	Pattern I
Type 1 diabetes; outcome of hepatitis V virus	<i>IL12B</i>	A/C SNP	See text	Pattern III
Cellular content of eosinophil cationic protein	<i>RNASE3</i>	562G>C polymorphism	Jonsson et al. (2006)	Pattern III
Familial porphyria cutanea tarda	<i>UROD</i>	G>A mutation	Cappellini et al. (2001)	Pattern I involving AU-rich sequence
Rheumatoid arthritis; breast cancer; sarcoidosis; renal disease	<i>VEGF</i>	936 C>T and 1451C>T polymorphisms in linkage disequilibrium	Krippel et al. (2003), Morohashi et al. (2003), Han et al. (2004), Doi et al. (2006)	936C>T: pattern of unknown significance, type 2; 1451C>T: pattern II involving a bulge loop (ACACC>ACACU)
Group 4: <i>Hitherto uncharacterised USS variants (n = 46)</i>				
Essential hypertension	<i>AGTRI</i>	A1166C polymorphism	Bonnardeaux et al. (1994), Wang et al. (1997), Gardier et al. (2004)	Pattern of unknown significance, type 1
Cardiovascular risk and open-angle glaucoma	<i>AGTR2</i>	3123C>A polymorphism	Deinum et al. (2001), Jones et al. (2003), Hashizume et al. (2005)	Pattern I involving AU-rich sequence
Colorectal cancer	<i>APC</i>	8636C>A polymorphism	Zhou et al. (2004)	Pattern II involving a loop comprising UUUAAAAGA(C/A)UGAAAAUUUUGUAAAUA
Familial adenomatous polyposis	<i>APC</i>	A8822G	Heinmann et al. (2001)	Pattern of unknown significance, type 2
Elevated triglyceride levels	<i>APOC3</i>	Polymorphic <i>Sst</i> I site (C>G)	Dammerman et al. (1993)	Pattern I
Human pigmentation	<i>ASIP</i>	g-8818A>G polymorphism	Kanetsky et al. (2002), Bonilla et al. (2005)	Pattern of unknown significance, type 1
Calcium oxalate stone disease; prostate cancer; primary open angle glaucoma	<i>CDHI</i>	3' UTR C/T polymorphism	Becker et al. (1995), Tsai et al. (2003), Wu et al. (2005), Lin et al. (2006)	Pattern II involving a loop comprising AGAAAUCA(C/U)GUUG
Human cancer	<i>CDKN1A</i>	C>T in 3' UTR	Mousses et al. (1995)	Pattern II involving a loop comprising AGCCUG(C/U)AG

Table 1 continued

Disease(s)/trait(s) under investigation ^a	Gene	Variant ^b	Main reference(s)	Predicted secondary structure change (and primary sequence or motif involved where appropriate) ^c
Melanoma and bladder cancer	<i>CDKN2A</i>	500C>G polymorphism	Kumar et al. (2001), Sakano et al. (2003)	Pattern III
Melanoma and bladder cancer	<i>CDKN2A</i>	540C>T polymorphism	Kumar et al. (2001), Sakano et al. (2003)	Pattern II involving a loop comprising (C/U)ACAACU
Sporadic pancreatitis	<i>CFTR</i>	c.4575+2G>A	Audrezet et al. (2002)	Pattern II involving a loop comprising A(G/A)AG
Major depression	<i>CHRM2</i>	A>T 1890 polymorphism	Comings et al. (2002)	Pattern I involving AU-rich sequence
Thyroid hormone metabolism	<i>DIO1</i>	C785T polymorphism	Peeters et al. (2003, 2005)	Pattern I involving AC-rich sequence
Thyroid hormone metabolism	<i>DIO1</i>	A1814G polymorphism	Peeters et al. (2003, 2005)	Pattern 0
Schizophrenia	<i>DPYSL2</i>	2236T>C polymorphism	Nakata et al. (2003), Hong et al. (2005)	Pattern of unknown significance, type 1
Essential hypertension	<i>EDN2</i>	A985G polymorphism	Sharma et al. (1999)	Pattern I involving mainly a GC-rich stem-loop structure
Blood pressure and essential hypertension	<i>ENDRA</i>	C1363T polymorphism	Nicaud et al. (1999), Benjafield et al. (2003)	Pattern of unknown significance, type 2
Haemophilia B	<i>F9</i>	A>G mutation	Vielhaber et al. (1993)	Pattern I
Blepharophimosis-ptosis-epicanthus inversus	<i>FOXL2</i>	nt 2293-2294 ins T mutation	Ojan et al. (2004)	Pattern I involving AU-rich sequence
Lipoxigenase metabolism	<i>GPX4</i>	T/C polymorphism	Villette et al. (2002)	Pattern I
Obesity	<i>IL6R</i>	rs2229238 (C/T) polymorphism	Wolford et al. (2003)	Pattern II involving a hairpin loop comprising GAUGA(C/U)AAAA
Type 2 diabetes	<i>KCNJ9</i>	8639A>G and 8999T>C in tight LD	Wolford et al. (2001)	8639A>G: pattern of unknown significance, type 2; 8999T>C: pattern I
Atopic dermatitis	<i>KLK7</i>	4 bp insertion (AACC) polymorphism	Vasilopoulos et al. (2004)	Pattern II involving a loop comprising AU-rich sequence
Blood pressure and carotid intima-medial thickness	<i>LEP</i>	Rare C538T variant	Gaukrodger et al. (2005)	Pattern of unknown significance, type 2
Atherosclerosis and Alzheimer's disease	<i>OLRI</i>	C>T SNP	Mango et al. (2003), Trabetti et al. (2006), Shi et al. (2006)	Pattern of unknown significance, type 2
Binge eating	<i>MC4R</i>	A1419G mutation	Branson et al. (2003)	Pattern II involving a loop comprising (A/G)CAU
Colorectal and bladder cancer	<i>NATI</i>	<i>NATI*10</i>	Gu et al. (2005), Chen et al. (2006)	Pattern I involving AU-rich sequence
Phenylketonuria	<i>PAH</i>	1378G>T mutation	Santana da Silva et al. (2003)	Pattern I involving UG-rich sequence
Increased risk of breast cancer	<i>PBB</i>	C>T polymorphism	Jupe et al. (2001), Spurdle et al. (2002), Campbell et al. (2003)	Pattern 0
Omphalocele	<i>PITX2</i>	delTAT mutation	Katz et al. (2004)	Pattern I involving AU-rich sequence
Waist-to-hip ratio; serum leptin levels in lean	<i>POMC</i>	C8246T polymorphism	Suviolahti et al. (2003), Baker et al. (2005)	Pattern I
Sporadic Creutzfeldt-Jakob disease	<i>PRND</i>	3' UTR+28 T>C polymorphism	Jeong et al. (2005)	Pattern I
Venous thrombosis	<i>PROCR</i>	4678C/G polymorphism	Medina et al. (2004, 2005)	Pattern II involving a loop comprising (C/G)UCAGAA
Plasma protein S levels	<i>PROS1</i>	C>A polymorphism at nt 2698	Leroy-Matheron et al. (1999)	Pattern I

Table 1 continued

Disease(s)/trait(s) under investigation ^a	Gene	Variant ^b	Main reference(s)	Predicted secondary structure change (and primary sequence or motif involved where appropriate) ^c
Carotid atherosclerosis	<i>PTGDS</i>	4111A>C polymorphism (rs6926)	Miwa et al. (2004)	Pattern II involving a loop comprising (A/C)CCCU
Colorectal cancer	<i>PTGS2</i>	PTGS2:9850; rs4648298 (A/G) polymorphism	Cox et al. (2004)	Pattern I
Non-small cell lung cancer	<i>PTGS2</i>	8473C>T polymorphism; rs5275	Campa et al. (2004)	Pattern I
Rheumatoid arthritis	<i>PTPN22</i>	C10442778T polymorphism	Carlton et al. (2005)	Pattern I
Hirschsprung disease	<i>RET</i>	3' UTR+124A>G polymorphism	Fitze et al. (2003)	Pattern I
Type 2 diabetes and hypertension	<i>RETN</i>	3' UTR+62G>A polymorphism	Tan et al. (2003)	Pattern II involving a loop comprising GG(G/A)GGA
Multiple sclerosis	<i>SPP1</i>	9583A/G polymorphism	Niino et al. (2003)	Pattern 0
Schizophrenia	<i>TCPI</i>	rs15982 (A/G) polymorphism	Yang et al. (2004)	Pattern of unknown significance, type 2
Myocardial infarction	<i>THPO</i>	A5713G polymorphism	Webb et al. (2001)	Pattern II involving a loop comprising C(A/G)UGU
Platelet count	<i>THPO</i>	rs6141(+24)G/A polymorphism	Garner et al. (2005)	Pattern I involving a GC-rich stem-loop structure
von Hippel-Lindau disease	<i>VHL</i>	852+10C>G	Klein et al. (2001)	Pattern I
Myocardial infarction	<i>ZNF627</i>	rs4804611	Shiffman et al. (2005)	Pattern III
Group 5: <i>Isolated examples</i> (<i>n</i> = 4)				
Heart failure	<i>ADRA2C</i>	21-bp insertion/deletion polymorphism	Small et al. (2004)	–
Asthma	<i>GPR44</i>	Two SNPs in LD	Huang et al. (2004)	–
mRNA instability	<i>TNFRSF1B</i>	Different combinations of three SNPs	Puga et al. (2005)	–
Fracture risk	<i>VDR</i>	Haplotype 1 comprising multiple SNPs	Fang et al. (2005)	–

Variants in groups 1 and 2 are arranged in the order in which they appear in the text, whereas those in the remaining groups are listed alphabetically in accordance with their gene symbols

^aSome of the disease- or trait-association data are controversial or should be regarded merely as putative until independently confirmed

^bNomenclature in accordance with the original publication; refer to Supplementary Tables S1–S5 in which all the variants have been annotated in the context of their genomic or mRNA reference sequences

^cSee both Table 4 and text for the definitions of the different secondary structural patterns

^dNot applicable or not done

^eSingle nucleotide polymorphism

Table 2 Summary of upstream sequence elements (USEs) that modulate cleavage/polyadenylation in human protein-coding genes

Gene	Gene symbol	Motif	Reference
Complement C2	<i>C2</i>	A ~50 bp tract of 42% U	Moreira et al. (1995, 1998)
Lamin B2	<i>LMNB2</i>	Multiple U (2–10) tracts	Brackenridge et al. (1997), Brackenridge and Proudfoot (2000)
Collagen, type I, $\alpha 1$	<i>COL1A1</i>	Multiple short U-rich tracts	Natalizio et al. (2002)
Collagen, type I, $\alpha 2$	<i>COL1A2</i>	Multiple short U-rich tracts	Natalizio et al. (2002)
Collagen, type II, $\alpha 1$	<i>COL2A1</i>	Multiple short U-rich tracts	Natalizio et al. (2002)
2'-5' oligo A synthetase	<i>OAS1</i>	A 25 bp tract of 40% U	Aissouni et al. (2002)
Cyclooxygenase-2	<i>PTGS2</i>	USE1, AUUUGAA; USE2, AUUUCUAA; USE3, AUUUCUUA	Hall-Pogar et al. (2005)
Prothrombin	<i>F2</i>	UUAUUUUUGUGUUU	Sachchithanathan et al. (2005)

The listed USEs represent those that have been experimentally confirmed. Recent bioinformatics analysis has, however, revealed that U-rich USEs are common in human genes (Legendre and Gautheret 2003; Hu et al. 2005).

An AU-rich element in the 3' UTRs of the *CCND1* and *ODC* genes is involved in eIF4E-mediated mRNA nuclear export

The eukaryotic translation initiation factor eIF4E promotes the nucleocytoplasmic export of cyclin D1 (*CCND1*) and ornithine decarboxylase (*ODC*) mRNAs but it does not influence *GAPDH* or actin mRNA export (Rousseau et al. 1996); this has recently been attributed to the presence of an eIF4E sensitivity element, a ~100 bp AU-rich sequence, in the USS of the human *CCND1* and *ODC* genes (Culjkovic et al. 2005).

Cis-acting elements that control mRNA subcellular localisation

In eukaryotic cells, localisation of mRNA to different regions of the cytoplasm ensures that intracellular proteins are synthesised near to where they are to function. It also provides the means to establish cellular polarity through asymmetric RNA and protein distribution (reviewed in Chabanon et al. 2004; Shav-Tal and Singer 2005). Table 3 summarises the currently well-characterised mRNA localisation signals (known as 'zipcodes') in mammalian genes.

3' UTR motifs that affect mRNA stability

Of the many known post-transcriptional gene regulatory mechanisms, the control of mRNA stability/instability is the most intensively studied.

AU-rich elements (AREs)

AU-rich elements were initially identified in the 3' UTRs of a variety of short-lived mRNAs encoding

cytokines and protooncogenes (Caput et al. 1986; Shaw and Kamen 1986). Recent bioinformatics analyses have, however, revealed a large repertoire of AU-rich mRNAs that encode functionally diverse proteins (Bakheet et al. 2001; Khabar et al. 2005). AREs from different mRNAs vary dramatically but can be divided broadly into three distinct classes based upon their sequence features: class I has 1–3 copies of scattered AUUUA motifs coupled with a nearby U-rich region or U stretch; class II has at least two overlapping copies of the nonamer UUAUUUA(U/A)(U/A) in a U-rich region; and class III has a U-rich region but lacks a core AUUUA element (Chen and Shyu 1995).

To date, at least a dozen ARE-binding proteins (AREBPs) have been identified (reviewed by Khabar 2005; Barreau et al. 2006) and these proteins can either *promote* or *inhibit* the degradation of ARE-containing mRNAs. In the former case, the exosome, a complex of 3' to 5' exoribonucleases, may be directly recruited to the ARE by certain AREBPs such as AUF1 (Chen et al. 2001); these AREBPs therefore function as mRNA-destabilising proteins. In the latter case, however, a putative DEXH RNA helicase, known as RHAU (RNA helicase associated with AU-rich element), may first be required to disrupt the interaction between ARE and certain AREBPs such as HuR before the exosome is activated (Tran et al. 2004); thus, these AREBPs function as mRNA-stabilising proteins.

mRNA-destabilising elements that are structurally different and functionally distinct from AREs have been claimed to be present in mRNAs encoding several cytokines including the granulocyte colony-stimulating factor (G-CSF; Brown et al. 1996; Putland et al. 2002) and tumour necrosis factor α (TNF- α ; Stoecklin et al. 2003). However, we believe that (a) these elements may well fall into the AU-rich element category

Table 3 Summary of 'zipcodes' in the USSs of mammalian genes

Gene (species)	Cells under investigation; localisation	Sequence element	<i>Trans</i> -acting protein(s) ^a	Reference
<i>c-myc</i> (mouse)	Fibroblasts; perinuclear	A single AU-rich stem-loop of 44 nucleotides	Annexin A2	Chabanon et al. (2005), Mickleburgh et al. (2005)
Slow troponin C (mouse)	Differentiated myocytes; perinuclear	A ~40-nucleotide-long region that forms multiple stem and loop structures	An uncharacterised myotube-specific 42-kDa polypeptide	Reddy et al. (2005)
Myelin basic protein (mouse)	Oligodendrocytes; myelin compartment	GCCAAGGAGCC contained within a stable stem-loop	Heterogeneous nuclear ribonucleoprotein A2	Ainger et al. (1997), Munro et al. (1999)
Alpha subunit of Ca ²⁺ /calmodulin-dependent protein kinase II (rat)	Cultured hippocampal neurons; dendrites	A region of 30 nucleotides containing a highly conserved C(G,C)CAGAGATCCCTCT motif in human, mouse, and rat	Not characterised	Mori et al. (2000)
Metallothionein-1 (rat)	Chinese hamster ovary cells; perinuclear	ACACCUCCACC contained within a complex secondary structure (the first CACC is predicted to be in a stem and the second CACC to form part of an internal loop)	Eukaryotic elongation factor 1 α	Mickleburgh et al. (2004), Nury et al. (2005)
Vimentin (human)	HeLa cells; perinuclear	Secondary structure	Eukaryotic elongation factor 1 α and HAX-1	Zehner et al. (1997), Al-Maghrebi et al. (2002)
β -actin (human)	H520, A549, and HT29 cell lines; cell periphery	ACACCC ^b	MLP1	Adereth et al. (2005)
Integrin α_3 (human)	H520, A549, and HT29 cell lines; cell periphery	ACACCC ^c	MLP1	Adereth et al. (2005)

Some of the elements identified in mouse and rat genes have been shown to be conserved in their human counterparts

^aDetected under the studied conditions

^bThe ACACCC motif in the chicken β -actin mRNA, located within the loop of a predicted hairpin structure, was found to bind to a 68-kDa protein (Ross et al. 1997). Our analysis by *mfold* (<http://www.bioinfo.rpi.edu/applications/mfold/old/rna/form1.cgi>) revealed that this motif in the human gene is also located within the loop of a predicted hairpin structure

^c*mfold* analysis revealed that the motif is involved in a loop with predicted complex secondary structure

primarily because they are highly AU-rich (i.e. CUGUUUAAUUAUUUAAACAG in Brown et al. 1996; UGUUUUCUGUGAAAAC in Stoecklin et al. 2003) and (b) that these elements may be regulated independently of other AREs in the same genes as a consequence of their interaction with different ARE-BPs. This notwithstanding, Brown et al. (1996) noted the importance of stem-loop structure in determining mRNA stability and suggested that it is the structure of the stem, rather than its sequence, which is important for function. We also evaluated the 15-nucleotide element reported by Stoecklin et al. (2003) and found that it could indeed form a stem-loop structure. Such AREs have been increasingly investigated in the context of secondary structure (e.g. Lopez de Silanes et al. 2004; Shao and Ismail-Beigi 2004; Berger et al. 2005; Fialcowitz et al. 2005).

C- or pyrimidine-rich elements

In contrast to the many short-lived mRNAs that often have half-lives measured in minutes, globin mRNAs, which have half-lives of >24 h, are among the most stable of mRNAs due to their obvious biological roles. Three discontinuous C-rich stretches (i.e. CCUGCC, GCCU, and CUCCCCUCCUUG) within the 3' UTR of α -globin mRNA have been identified as being essential for the longevity of the mRNA in erythroid cells (Weiss and Liebhaber 1994, 1995). They bind to α CP1 and α CP2 [α -globin poly(C)-binding proteins; Kiledjian et al. (1995)] which also interact directly with poly(A)-binding protein (PABP; Wang et al. 1999). This complex RNA-protein-protein interaction may promote/stabilise the binding of PABP to the poly(A) tail, thereby hindering the deadenylation of the α -globin mRNA (Wang et al. 1999). Interestingly, all three α CP-binding, C-rich segments are located in the loop or interloops of a complex stem-loop structure (Waggoner and Liebhaber 2003).

C-rich elements have also been found in the 3' UTRs of three other highly stable mRNAs: α (I)collagen, tyrosine hydroxylase and 15-lipoxygenase. That all these C-rich elements bind to one or more common *trans*-acting protein factors may suggest a common regulatory mechanism of mRNA stability (reviewed by Waggoner and Liebhaber 2003). These C-rich elements are also known as pyrimidine-rich elements since the second frequently encountered nucleotide within the element is U (e.g. Holcik and Liebhaber 1997). In this regard, Yu and Russell (2001) identified a specific 14-nt pyrimidine-rich

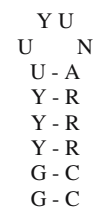
tract, UUCCUUUGUCCCCU, in the 3' UTR of the β -globin mRNA as the stability determinant. Given that antibodies raised against α CP also bind to the β -globin mRNP complex, these authors suggested that the highly stable α - and β -globin mRNAs might be regulated through a common pathway.

AG-rich elements

In keeping with studies of the orthologous murine gene (Glisovic et al. 2003 and references therein), Christian et al. (2004) demonstrated that the stability of the human *CYP2A6* mRNA is increased when its 3' UTR is bound by heterogeneous nuclear ribonucleoprotein (hnRNP) A1. Further, secondary structure prediction studies revealed that the primary binding site within the 3' UTR of the *CYP2A6* mRNA contains AG-rich blocks that are likely to be single-stranded (Christian et al. 2004).

Histone mRNA 3'-terminal stem-loop

The half-life of replication-dependent histone mRNA is about 40 min during S phase but decreases to 10 min by the end of S phase. This rapid degradation appears to be due to the release of HBP (hairpin-binding protein) from a 6-bp stem/4-bp loop structure



(Y=C/U; R=G/A; N=A/G/C/U) in the gene's 3' UTR, allowing 3' exonuclease to access the histone mRNA (reviewed in Jaeger et al. 2005).

Cis-acting elements determining site-specific endonucleolytic mRNA cleavage

The degradation of at least four human mRNAs—Gro protein alpha (*groa* or *CXCL1*; Stoeckle 1992), transferrin receptor (*TFRC*; Binder et al. 1994), insulin-like growth factor 2 (*IGF2*; de Pagter-Holthuizen et al. 1988), and α -globin (Wang and Kiledjian 2000)—is, at least in part, determined by site-specific endonucleolytic cleavage within their 3' UTRs. The frequency of occurrence of this regulatory mechanism is, however, likely to be seriously underestimated because

internally cleaved mRNA intermediates are difficult to capture: whilst the resulting 5' fragment is no longer protected by a poly(A) tail (analogous to deadenylation), the resulting 3' fragment is unprotected by a cap structure (analogous to decapping; Mata et al. 2005). The mRNA breakdown intermediates captured in the above four cited cases may have formed stable RNase-resistant duplex structures or could have bound to stabilising *trans*-acting factors (Ross 1995). To date, endonucleolytic cleavage sites (ECS) have been fully characterised in three of the four mRNAs.

The role of iron-responsive elements (IREs) in modulating TFRC mRNA stability

TFRC mRNA displays a relatively short half-life when iron is abundant and a relatively long half-life when iron is scarce (Koeller et al. 1991). Regulation is accomplished through the binding of *trans*-acting IRE-binding proteins (IRPs) to the *cis*-IREs located within the gene's 3' UTR (Casey et al. 1989; Koeller et al. 1989).

The iron-responsive element is a hairpin structure comprising a 5'-CAGWGH-3' apical loop (W is A or U and H is C, A or U; underlined bases form a Watson-Crick pair) and a stem that is interrupted either by a single C bulge or an unpaired C residue within an internal loop/bulge. The apical loop is necessary for high affinity binding to IRPs, whereas the C-bulge appears to orientate optimal protein binding without directly contacting the IRPs (Meehan and Connell 2001).

Up to five IREs (viz. A, B, C, D, and E in the order 5' to 3') have been identified in the 3' UTR of the *TFRC* gene (Casey et al. 1988); three (i.e. B, C, and D) were found to be capable of conferring iron-independent regulation of *TFRC* expression upon binding with IRPs (Casey et al. 1989). How does the IRE-IRP association stabilise *TFRC* mRNA? Binder et al. (1994) detected a shorter mRNA that lacks a significant portion of the 3' end of the *TFRC* mRNA in a human plasmacytoma cell line. These authors noted that (a) the appearance of the truncated molecules correlated with rapid turnover of the *TFRC* mRNA; (b) the truncated RNA resulted from endonucleolytic cleavage rather than a 3' → 5' exonuclease pause; and (c) the ECS was located just 6 bp 3' to the IRE element C, and concluded that the binding of IREs with IRPs results in protection of the *TFRC* mRNA against nucleolytic attack. Put another way, the IREs and the ECS constitute the key determinants of *TFRC* mRNA stability.

Secondary structures formed by two widely separated sequence elements are required for the specific cleavage of IGF2 mRNA

Human *IGF2* mRNA is specifically cleaved between nucleotides 2183 and 2184 (the first nucleotide 3' to the translational stop codon as 1; Supplementary Table S3) in the 3' UTR (de Pagter-Holthuizen et al. 1988; Meinsma et al. 1991). The ECS, located in an internal loop, is flanked by two complex secondary structures: the 5' flanking one contains two stable stem-loops that are formed by nucleotides -133 to -7 (numbered with respect to the ECS); the 3' flanking one comprises a downstream G-rich tract (element II) that folds into an extended duplex of 83 nucleotides with a C-rich sequence tract (element I) that is located ~2 kb 5' to the ECS (Meinsma et al. 1992; Scheper et al. 1995; van Dijk et al. 1998, 2000). These secondary structures (refer to Fig. 1 in van Dijk et al. 2000) have been shown to maintain a highly specific ECS by preventing the formation of alternative structures.

C-rich elements and specific α -globin mRNA cleavage

Three C-rich elements in the 3' UTR of α -globin mRNA have been postulated to hinder deadenylation through the formation of an RNA- α CP-PABP complex. α -globin mRNA has also been reported to be specifically cleaved in its 3' UTR, the ECS (indicated by /) being located within the third and longest C-rich element, CUCCCCUCC/UUG (Wang and Kiledjian 2000). More interestingly, secondary structure prediction has suggested that whilst the first nine nucleotides of the element are single stranded, the last three nucleotides are located within a stem structure. In other words, the ECS is located just 5' to a double-stranded region (Waggoner and Liebhaber 2003). This degradation pathway is probably also regulated by the interaction between the C-rich elements and α CPs, irrespective of whether it is functionally linked with the deadenylation pathway (Waggoner and Liebhaber 2003).

Cis-acting 3'UTR elements that control mRNA translation

AU-rich elements

AU-rich elements (AREs) are the best-known determinants of mRNA stability/instability. However, it has been known for some time that AREs within certain cytokine genes are also responsible for translational

repression (effects on mRNA stability have been excluded; Krays et al. 1989; Han et al. 1990). The expression of some 14 genes, mostly encoding proteins involved in inflammation and tumour growth, is known to be controlled at the level of translation by means of an association between AREs and AREBPs such as TIA-1 and CUGBP2 (Espel 2005).

T-cell intracellular antigen 1 (TIA-1) has been best characterised as a translational repressor in the context of the ARE-bearing tumour necrosis factor- α (*TNFA*) and cyclooxygenase 2 (*COX2*) genes (Piecyk et al. 2000; Dixon et al. 2003). Very recently, using immunoprecipitation of TIA-1–RNA complexes from human colorectal carcinoma RKO cells followed by microarray-based identification and computational analysis of bound transcripts, Lopez de Silanes et al. (2005) identified a common motif present in TIA-1 target mRNAs. The 30 to 37-nucleotide-long motif is highly U-rich in its 5' segment and AU-rich in its 3' segment, forming loops of variable size and a bent stem. This motif has not only been found in both the *TNFA* and *COX2* mRNAs but also in ~3% (3,019) of UniGene transcripts (Lopez de Silanes et al. 2005).

Cis-acting sequence involved in specific incorporation of selenocysteine at UGA codons of selenoprotein mRNAs

In addition to functioning as a translational stop codon, UGA also signals the incorporation of selenocysteine (Sec). In eukarya and archaea, recognition of UGA as Sec requires a specific secondary structure known as the selenocysteine-insertion sequence (SECIS) that is located within the USS of these mRNAs. The SECIS recruits SECIS-binding protein 2 (SBP2), which in turn recruits a Sec-specific translation elongation factor and a specific Sec transfer RNA (Berry 2005 and references therein).

The canonical mammalian SECIS element is characterised by a hairpin structure comprising two helices, an internal loop, four consecutive non-Watson-Crick base-pairs (the quartet) containing a central G.A/A.G tandem, and an apical loop (Walczak et al. 1996). [Unlike 'Watson-Crick base-pairs', 'non-Watson-Crick base-pairs' present chemical groups in either the major or the minor groove of the helix that are available for specific protein or RNA binding (Walczak et al. 1998) and may thus be regarded as single stranded]. However, more recent studies have led to the identification of non-canonical forms of SECIS (Korotkov et al. 2002; Kryukov et al. 2003). Indeed, as Korotkov et al. (2002) opined, the absolutely conserved primary

sequence in the SECIS is limited to the UGA...GA motif in the quartet, which serves as a specific recognition site for SBP2; the only other recognition feature of the SECIS might be its three-dimensional structure.

MicroRNA targets: another type of *cis*-acting regulatory element

MicroRNAs (miRNAs) post-transcriptionally regulate gene expression by binding to their target mRNAs (usually within the 3' UTRs) (for recent reviews, see Bartel 2004; Pillai 2005; Valencia-Sanchez et al. 2006). The total number of human miRNAs has recently been estimated to be at least 800 (Bentwich et al. 2005), each being potentially capable of down-regulating a large number of different target mRNAs (Lim et al. 2005). However, to date, the mechanism by which the bound miRNA down-regulates gene expression remains unclear. In this regard, it is pertinent to note that Robins et al. (2005) and Zhao et al. (2005) have recently tried to incorporate the folded structure of mRNA to understand miRNA–target interactions.

Practical considerations about secondary structure prediction for establishing the rules that determine how RNA regulatory elements function in the context of a specific secondary structure

The above fairly comprehensive survey of well-defined *cis*-regulatory elements within the 3' UTRs of protein-coding genes supports the view that RNA regulatory elements function in the context of a specific secondary structure, best exemplified by the extensively studied histone mRNA 3'-terminal stem-loop, IRES, SECIS, and *cis*-elements determining site-specific endonucleolytic mRNA cleavage. To establish general rules governing how RNA secondary structure might operate to effect the functionality of component regulatory elements, the most commonly used *mfold* program (Zuker 2003; <http://www.bioinfo.rpi.edu/applications/mfold/old/rna/form1.cgi>) was employed to predict secondary structures under default parameters. Several key points regarding this analysis are discussed below.

'Local' vs 'global'

Our analysis was based on the prediction of 'local' rather than 'global' secondary structure. Interested readers are invited to consult Meyer and Miklos (2005) for detailed discussions of this issue.

Sequence length used for secondary structure analysis

Secondary structure varies as a function of the input sequence length for which there is no consensus for use in *mfold* analysis [e.g. 36 bp in Ruggiero et al. (2003); 87 bp in Chabanon et al. (2005); from 100 to 200 bp in Aranda-Abreu et al. (2005); 130 bp in Waggoner and Liebhaber (2003); 234 bp in Christian et al. (2004)]. In an attempt to evaluate the diverse USS variants in both a systematic and an objective manner, we employed the following policy: (a) the sequence analysed was always limited to the last exon; (b) ± 100 bp sequences flanking each variant were used for analysis wherever possible; (c) if either of the analysed paired sequences (i.e. wild-type and variant sequences) did not display an obvious folded structure or the 5' flanking sequence was < 100 bp, ± 50 bp sequences flanking each variant were used for analysis wherever possible; and (d) if again either of the paired sequences analysed did not display a folded structure, ± 60 or ± 40 bp sequences flanking each variant were used for analysis wherever possible.

Terms used for describing secondary structural features

Terms used for describing RNA secondary structural features (e.g. hairpin loop, bulge loop, internal loop) are in accordance with Mathews et al. (1999).

Evaluating non-functional USS variants in the context of secondary structure

Four USS variants were identified as being non-functional during the course of this meta-analytical study.

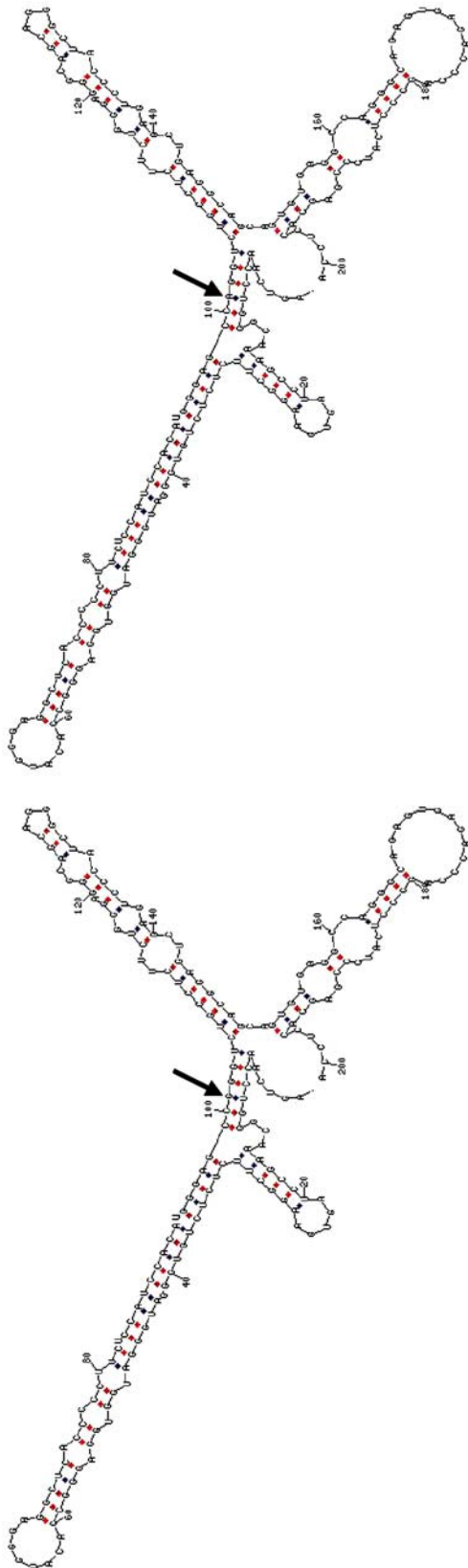
CXCL12 Stromal cell-derived factor-1 (SDF-1; or CXCL12; MIM# 600835) is the principal ligand for CXCR4, a co-receptor with CD4 for T-lymphocyte cell line-tropic human immunodeficiency virus type 1 (HIV-1). Winkler et al. (1998) identified a common polymorphism, designated SDF1-3'A (or SDF-G801A) in the 3'-UTR of the SDF-1 β isoform. Homozygosity for SDF1-3'A has been shown in one study to have a protective effect with respect to AIDS progression (Winkler et al. 1998) although it has also been reported to be associated with accelerated disease progression (Mummidi et al. 1998; van Rij et al. 1998). A subsequent international meta-analysis has, however, indicated that SDF1-3'A homozygosity is unlikely to affect disease progression (Ioannidis et al. 2001).

If the *SDF1-3'A* allele really were to play a role in AIDS progression, then it should be associated with altered gene expression. In the original work of Winkler et al. (1998), the functional significance of the *SDF1-3'A* allele was simply inferred from the observation that “the *SDF1-3'A* variant is located in a segment of the 3'-UTR of the *SDF-1 β* transcript that is highly conserved in sequence (69% sequence between human and mouse *SDF-1 β* 3' UTR sequence with no gaps).” As opined by Winkler et al. (1998), “this extent of conservation with the segment suggests that it may serve as a target for *cis*-acting factors influencing transcript abundance, synthesis, transport, stability, or splice product abundance.” However, the *SDF1-3'A* variant was not found to affect *SDF-1 β* RNA synthesis either in vitro (Arya et al. 1999) or in vivo (Kimura et al. 2003). More recently, using allele-specific transcript quantification, Kimura et al. (2005) were able to show that polymorphisms other than *SDF1-3'A* exert a *cis*-acting effect on the expression of *SDF-1* transcripts.

APOA5 Apolipoprotein A-V (*APOA5*; MIM# 606368) plays an important role in the regulation of triglyceride metabolism (Merkel and Heeren 2005). The *APOA5*2* haplotype, which contains a T>C SNP in the 3' UTR (located 158 bp downstream of the translational termination codon) and is present in ~16% of Caucasians, was found to be significantly associated with increased plasma triglyceride levels (Pennacchio et al. 2001, 2002). The 3' UTR T>C SNP has, however, recently been shown not to influence reporter gene expression (Talmud et al. 2005).

NPPC A rare G>A polymorphism in the C-type natriuretic peptide gene (*NPPC*; MIM# 600296) has been reported to be tentatively associated with hypertension in the Japanese population (A allele frequency: 1.6% in controls vs 2.6% in patients). However, this polymorphism was not found to influence luciferase activity when evaluated in a transient reporter gene assay (Ono et al. 2002).

SEPI5 *SEPI5*, one of the 25 selenoprotein-encoding genes (Kryukov et al. 2003), manifests two haplotypes, C811/G1125 versus T811/A1125 (frequencies: 68 vs 32%); both variants are located within the USS of the gene's 3' UTR (Gladyshev et al. 1998; Kumaraswamy et al. 2000). Unlike G>A at position 1125 (which will be considered in the next section), C>T at position 811 was not predicted to reside within a SECIS motif using SECISearch 2.19 (<http://www.genome.unl.edu/SECISearch.html>; Kryukov et al. 2003); C811T has also been consistently found to be non-functional in a reporter



◀ **Fig. 1** Proposed secondary structures of the wild-type (*left panel*) and variant (*right panel*) RNA sequences in relation to the alternative G/A alleles (position 101; indicated by *arrow*) of the 3' UTR polymorphism in the *CXCL12* gene. This type of secondary structural change was termed pattern 0

gene system (Kumaraswamy et al. 2000; Hu et al. 2001).

All four non-functional USS variants result in a similar secondary structural change

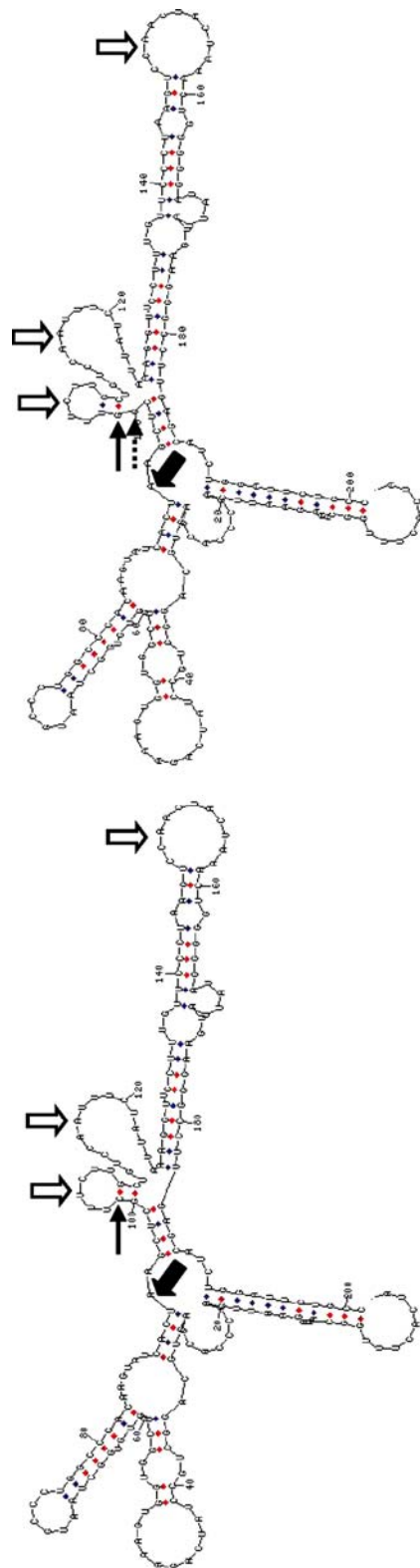
All four non-functional USS polymorphisms were predicted to give rise to similar secondary structures (here termed pattern 0) in which both the substituted and substituting nucleotides are capable of pairing with the same opposing nucleotides in a helical portion (NB. G can pair with either U or C; U can pair with either G or A in an mRNA molecule). This is exemplified by the G801A polymorphism in *CXCL12* (Fig. 1); the predicted secondary structures of the other three polymorphisms are provided in Supplementary Figs. S1, S2 and S3, respectively.

Evaluating functional USS variants in the context of secondary structure

A total of 17 known functional USS variants were collated and analysed individually. All deletions of ≥ 5 bp were predicted by *mfold* to result in significant secondary structural change and will thus not be discussed individually in the text. As we shall see below, the functional consequences of these variants, taken together, affected virtually all the main stages of post-transcriptional gene regulation (with the exception of subcellular localisation). In addition, the *cis*-regulatory elements that these variants have disrupted can also be confidently assigned in most cases.

HBB: a point mutation occurring within a CU-rich sequence tract affects both mRNA 3' end formation and stability

A C>G mutation, six nucleotides downstream of the translational termination codon (UAA) of the *HBB* gene (MIM# 141900; encoding β -globin), has been reported in β -thalassaemia intermedia patients (Jankovic et al. 1991; Maragoudaki et al. 1998). This mutation, which occurred within a pyrimidine-rich sequence



◀ **Fig. 2** Proposed secondary structures of the wild-type (*left panel*) and mutant (*right panel*) RNA sequences in relation to the 3' UTR C>G mutation (position 101; indicated by *thin solid arrow*) in the *HBB* gene. *Dotted arrow* the unpaired G residue in the mutant sequence; *thick solid arrow* translational stop codon UAA (positions 93–95); *open arrows* loops comprising pyrimidine-rich sequences. This type of secondary structural change was termed pattern I

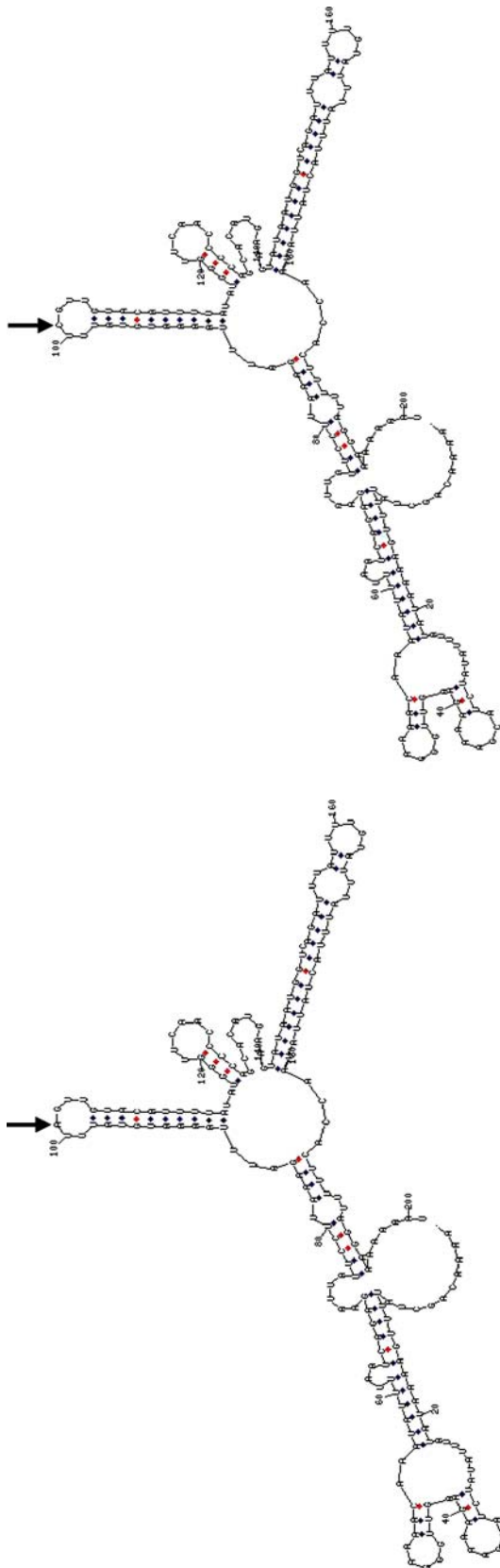
1998). Thus, in vivo, the mutant allele is associated with a 40–68% reduction in *HBB* mRNA level as compared with the wild-type allele. This observation concurs with data obtained from in vitro analysis: steady-state cytoplasmic mRNA levels from transfected MEL cells containing the mutant allele were reduced by 52–60% as compared with those obtained from the wild-type allele (Sgourou et al. 2002). Analysis of nuclear RNA demonstrated that the *HBB* mutation served to lower the ratio of cleaved/uncleaved transcripts by 22–30%, suggesting that the C>G substitution adversely affects mRNA 3' end formation. As already noted by Sgourou et al. (2002), the observed decrease in nuclear *HBB* RNA does not account for the rather greater decrease in cytoplasmic *HBB* mRNA. It may well be therefore that this lesion impacts on other properties of the *HBB* mRNA such as its stability.

As shown in Fig. 2, all three large loops that formed downstream of the translational stop codon (UAA) are pyrimidine- or U-rich. Since U-rich elements are known to regulate mRNA 3' end formation (refer to Table 2), and CU-rich elements in the 3' UTR of *HBB* mRNA have been shown to affect mRNA stability (Yu and Russell 2001), it is reasonable to suppose that these CU-rich sequence tracts regulate both pre-mRNA 3' end formation in the nucleus and mRNA stability in the cytoplasm. Not surprisingly, the C>G mutation was predicted not only to lead to a shortening of the first bulge loop by one base (i.e. from UUUCUU to UUUCU) but also to introduce an unpaired base (G) into the stem portion of this loop (Fig. 2). Thus, whilst the shortened CU-rich motif may exhibit reduced binding affinity for *trans*-acting factors, the unpaired G within the stem could also have affected the interaction between the CU-rich motif(s) and the *trans*-acting factors by modifying the orientation of the first bulge loop. This kind of secondary structure change, affecting both the loop(s) and stem(s), was termed pattern I.

HBB: a 13 bp deletion that may affect RNA nucleocytoplasmic transport

tract, UAAGCUCG(C/G)UUUCUUGCUGUCCAAUUUCUAUU, is associated with a 20–34% reduction in *HBB* mRNA levels in heterozygous patients as compared with healthy controls (Maragoudaki et al.

A 13 bp deletion (GCATCTGGATTCTGCCTAA-TAAA) that terminates four nucleotides 5' to the UCPAS (in italics) of the *HBB* gene was identified in a



◀ **Fig. 3** Proposed secondary structures of the wild-type (*left panel*) and variant (*right panel*) RNA sequences in relation to the 3' UTR A>C SNP (position 101; indicated by *arrow*) in the *BMP2* gene. This kind of secondary structural change was termed pattern II

12-year-old Turkish boy with β -thalassaemia (Basak et al. 1993). This mutation was associated with a 6-fold reduction in mRNA level in vivo. However, in vitro analysis indicated that the lesion neither affected the assembly of the mRNA-stabilising mRNP complex nor the stability of the mutant mRNA. Rather, this deletion appears to decrease *HBB* mRNA levels mainly by inhibiting the processing of pre-mRNA to mRNA in the nucleus (Bilenoglu et al. 2002). The original authors considered that the mutation might disrupt a sequence element that facilitates nucleocytoplasmic transport, thereby hampering the processing of the mutant pre-mRNA, and resulting in the accumulation of the fully processed RNA species in the nucleus.

USS variants that affect mRNA stability

AU- or U-rich elements

Six USS variants involving six different genes were found to affect AU- or U-rich elements.

BMP2. Variations in the *BMP2* gene (MIM# 112261; encoding bone morphogenetic protein 2) have been associated with both osteoporosis (Styrkarsdottir et al. 2003) and osteoarthritis (Valdes et al. 2004). A common A>C SNP in *BMP2*, located within the AU-rich USS, was found to be functional: the minor C allele not only has a different affinity for specific proteins but also exhibits a higher in vitro decay rate as compared with the major A allele (Fritz et al. 2006). Unlike the above-mentioned pattern 0 and pattern I variants, the A>C SNP in the *BMP2* gene was located within a hairpin loop (here termed pattern II; Fig. 3). Since this was the only change observed in the predicted secondary structure, it would follow that an A to C substitution in a single-stranded AU-rich motif might reduce the motif's affinity for its cognate binding factors.

NR3C1. The human glucocorticoid receptor gene (*GCCR* or *NR3C1*; MIM# 138040) comprises 10 exons (1–8, 9 α , and 9 β). It encodes two major isoforms through alternative splicing; whereas the α isoform (hGR α ; exons 1–9 α) encodes a functional receptor, the β isoform (hGR β ; exons 1–8 and 9 β) encodes a protein that does not bind glucocorticoid (Oakley et al. 1996).

However, hGR β acts as a dominant negative inhibitor of hGR α by competitively binding to hGR-interacting proteins (Oakley et al. 1996; Charmandari et al. 2005). Thus, increased expression of hGR β would be predicted to result in glucocorticoid resistance which may not only impede glucocorticoid treatment of immune-related disease but may also contribute to the pathogenesis of these conditions (Schaaf and Cidlowski 2002).

An A>G polymorphism, which occurred in the AU-rich USS of *hGR β* mRNA, has been reported to be associated with rheumatoid arthritis (Derijk et al. 2001). Two independent in vitro studies have shown that the G allele of this polymorphism increases the stability of the reporter mRNA (Derijk et al. 2001; Schaaf and Cidlowski 2002). This polymorphism also results in a pattern II secondary structural change involving an AU-rich loop (Supplementary Fig. S4).

TGFB3. Beffagna et al. (2005) recently detected a 1723 C>T transition in the 3' UTR of the transforming growth factor- β 3 gene (*TGFB3*; MIM# 190230) in a 16-year-old boy with a typical arrhythmogenic right ventricular dysplasia (ARVD) phenotype and his brother who died suddenly of ARVD at the age of 16. This variant was considered to be disease-associated since it was not found in 600 control chromosomes, and an in vitro transfection assay revealed that it significantly increased reporter gene activity in murine C2C12 myoblasts (Beffagna et al. 2005).

The *TGFB3* C>T polymorphism is again predicted to result in a pattern II secondary structural change involving an AU-rich sequence (Supplementary Fig. S5). However, while the above-mentioned *BMP2* and *NR3C1* polymorphisms decrease the number of motif-defining nucleotides, the *TGFB3* polymorphism increases the number of such nucleotides.

CEACAM1. A single T deletion polymorphism, located within a T8 sequence tract in the 3' UTR of the carcinoembryonic antigen-related cell adhesion molecule 1 gene (*CEACAM1*; MIM# 109770), is involved in tumour onset and progression; this Δ T allele is also associated with an increased level of reporter gene expression (Ruggiero et al. 2003). This deletional polymorphism has been predicted to result in an enlarged T-rich hairpin (from 4T to 6T), a shortened stem, and the generation of a new small bulge loop in the deletion allele as compared with the insertion allele (see Fig. 2 in Ruggiero et al. 2003). This then falls into the pattern I secondary structure change.

PPP1R3A. A common polymorphism comprising a 5 bp deletion plus three single nucleotide substitutions in the AU-rich 3' UTR of the protein phosphatase-1 regulatory subunit 3 gene (*PPP1R3A*; MIM# 600917) has been found to be associated with insulin resistance and type 2 diabetes in Pima Indians; it also correlates with significantly reduced *PPP1R3A* mRNA expression in vivo (Xia et al. 1998). As illustrated in Fig. 4, the deletion allele is characterised by the loss of an AU-UUA motif and the generation of a similar motif, AUUUUA; the spacing of the two AUUU(U)A motifs nevertheless differs between the two alleles. Transient expression analysis indicated that the half-life of the chimeric β -globin mRNA containing the deletion allele was at least 10-fold shorter than that observed for an mRNA species bearing the insertion allele, suggesting that the observed reduction in *PPP1R3A* mRNA could be directly attributable to the deletion variant in the 3' UTR (Xia et al. 1998). Further in vitro analyses revealed that three different proteins (43, 80, and 139 kDa) bind to the polymorphic region whereas the less stable deletion allele exhibits \geq 2-fold higher relative protein binding (Xia et al. 1999).

TYMS. A common 6-bp deletion polymorphic variant (delTTAAAG) in the USS of the thymidylate synthase gene (*TYMS*; MIM# 188350; Ulrich et al. 2000) has recently been reported to be associated with gastric (Graziano et al. 2004; Zhang et al. 2005) and colorectal (Mandola et al. 2004) cancer susceptibility, sensitivity of gastric cancer to 5-fluorouracil-based chemotherapy (Lu et al. 2006), and clinical outcome of patients with esophageal adenocarcinoma treated with pre-operative chemoprevention (Liao et al. 2006). That this deletion variant is functional is supported by two observations: (1) it is associated with reduced intratumoural *TYMS* mRNA expression in vivo and (2) the \sim 35% decrease in the in vitro expression of *TYMS*-3' UTR reporter gene constructs bearing the 6 bp deletion was due to an increased rate of mRNA degradation (Mandola et al. 2004).

C-rich element in *NPRI*

A common micro-deletion (CCCC \rightarrow CCC) polymorphism (with an allele frequency of 33–39% in the general population) in the 3' UTR of the human natriuretic peptide receptor A gene (*NPRI*;

5' aguaauugaacaugaaauuuguuuuuaacacuguaucuuuuaucaggauucaugau3' Insertion allele
5' aguaauugaacaugaaauuugu-----uuuuuaucuuuuaucaggauucaugau3' Deletion allele

Fig. 4 Comparison of the deletion and insertion alleles of the length polymorphism in the 3' UTR of the *PPP1R3A* gene. The

AUUU(U)A motif is shaded. Dashes indicate the deleted residues. The three substitutions are highlighted in bold

MIM#108960) was identified concurrently by Knowles et al. (2003) and Pitzalis et al. (2003). These two independent studies were complementary to each other. Transient expression analysis of constructs containing the *NPRI* 3' UTR indicated that the 3C allele was associated with a 3-fold reduction in the expression of the reporter gene by comparison with the 4C allele (Knowles et al. 2003). On the other hand, although the 3C allele is not associated with cardiovascular disease, it does appear to represent a risk-modifying factor: the 3C allele, in heterozygous form, occurs more frequently in young normotensive subjects with a family history of hypertension than those without a family history, whereas 3C homozygotes display significantly higher systolic blood pressure and a prolonged ventricular relaxation time by comparison with 4C homozygotes (Pitzalis et al. 2003).

In the words of Knowles et al. (2003), “we were surprised to find that having 3C or 4C at 30 bp downstream of the stop codon in the *NPRI* transcript significantly affected expression”. With hindsight, however, this finding is compatible with the observation that C-rich elements act as mRNA stability determinants (Waggoner and Liebhaber 2003). Further, the deletional polymorphism (underlined) is located within an extended C-rich sequence tract, CCTG CCTCTCTCTATCCCTCCACACCTCCC(C)TAC CC with most of the cytosines being predicted to be single-stranded (Supplementary Fig. S6). The loss of one cytosine in the hairpin loop of the deletion allele (pattern II) could reduce the binding efficiency to mRNA-stabilising *trans*-acting factors.

CU-rich element in TFCP2

Lambert et al. (2000) reported that the A allele of a G/A polymorphism in the 3' UTR (15 bp downstream of the translational stop codon) of the transcription factor LBP-1c/CP2/LSF gene (*TFCP2*; MIM# 189889) was associated with a reduced risk of Alzheimer's disease (AD) in three independent populations (French, British, and North American). Although two studies (Taylor et al. 2001; Luedeking-Zimmer et al. 2003) have lent support to this finding, other studies have yielded contradictory results viz. the A allele was found to be associated with increased AD risk (Panza et al. 2004; Bertram et al. 2005). The underlying reasons for this discrepancy are unclear but it may be that the 3' UTR SNP is not itself pathogenic but is rather in linkage disequilibrium (LD) with another functional genetic variant in the vicinity. This notwithstanding, electrophoretic mobility shift assays have demonstrated that oligomers containing the A

allele display, on average, a 3.75-fold lower affinity for neuroblastoma nuclear proteins than those containing the G allele, whereas the absence of the A allele was associated with lower expression of *TFCP2* mRNA. These findings suggested that *TFCP2* gene expression may be differentially modulated by the G and A alleles within the 3' UTR sequence through differential binding of a nuclear protein (Lambert et al. 2000).

The G/A polymorphism, occurred within a CU-rich sequence tract, CGUUUC(G/A)UGCCC, was predicted to cause a pattern I secondary structural change (Supplementary Fig. S7). In particular, this polymorphism significantly affects the length of a hairpin loop: although the wild-type hairpin loop contains 12 CU-rich bases, the variant loop contains only three bases (UUC). In addition, the newly generated 5-nt bulge loop in the variant allele is not CU-rich (Supplementary Fig. S7). In summary, the G>A polymorphic change may convert much of the normal CU-rich motif from the single-stranded state to the double-stranded state, thereby lowering its binding affinity for *trans*-acting factors. This structural analysis therefore argues strongly for the functional significance of the G>A polymorphic change, although its exact role in the pathogenesis of AD remains to be clarified.

AG-rich elements

In addition to the recently described functional polymorphisms that have occurred within different AG-rich blocks in the 3' UTR of the *CYP2A6* gene (Wang et al. 2006), three other polymorphisms were retrospectively found to occur within an AG-rich element.

F7. Factor VII (*F7*; MIM# 227500) plays a key role in the intrinsic pathway of blood coagulation. A 2 bp (AA) insertional polymorphism in the 3' UTR of the *F7* gene, with an allele frequency of 15% in the Caucasian population, is often associated with other variants causing *F7* deficiency (Peyvandi et al. 2005 and references therein). It has been postulated that this common polymorphism may in part account for the poor genotype–phenotype correlation that has frequently been observed in *F7* deficiency (Peyvandi et al. 2005). Transient expression analysis has demonstrated that the insertion allele reduces the steady-state level of *F7* mRNA by 40% as compared with the deletion allele (Peyvandi et al. 2005).

The AA insertional polymorphism in the *F7* gene occurs within a large loop of AG-rich sequence tract and results in a pattern II secondary structure change (Supplementary Fig. S8). The consequently enlarged single-stranded AG-rich motif might display an increased binding capacity for *trans*-acting factors.

PTPNI. Di Paola et al. (2002) identified a common G micro-insertion variant at position 1484 (nomenclature in accordance with GenBank accession number M33689) in the 3' UTR of the protein tyrosine phosphatase 1B gene (*PTPNI*; MIM# 176885) that is associated with several features of insulin resistance in two different Italian populations. Although this association has not been confirmed by subsequent studies (Echwald et al. 2002; Dahlman et al. 2004; Florez et al. 2005; Spencer-Jones et al. 2005), the potential functionality of this insertion polymorphism is indicated by (a) *PTPNI* mRNA levels being significantly higher in five muscle biopsies taken from 1484insG carriers than in 11 age- and sex-matched controls and (b) 1484insG mRNA being more stable than that derived from the deletion allele in transfection studies (Di Paola et al. 2002). This deletion/insertion polymorphism occurs within an AG-rich sequence tract and results in a pattern I secondary structural change (Supplementary Fig. S9).

NPRI. In addition to the common micro-deletion (CCCC → CCC) polymorphism, Knowles et al. (2003) also identified a second 4 bp (AGAA) micro-deletion polymorphism in the 3' UTR of the *NPRI* gene, with an allele frequency of 4%. Transient expression analysis has shown it to be associated with a 2-fold reduction in expression of the reporter gene (Knowles et al. 2003). This deletion polymorphism, which occurs within an AG-rich sequence tract, GGGAGGAGAAAGAG, results in a pattern I secondary structural change (Supplementary Fig. S10). Thus, the stability of *NPRI* mRNA appears to be regulated by two different kinds of 3' UTR motifs (i.e. C-rich and AG-rich elements).

Variants that modify the efficiency of Sec incorporation

Of the 25 known selenoprotein-coding genes, four have been reported to harbour disease-associated variants in their 3' UTRs. However, only the T>C mutation in *SENPI* (MIM# 606210), the G1125A polymorphism in *SEPI5* (MIM# 606254) and a common T/C polymorphism in *GPX4* (MIM# 138322) appear to reside within the SECIS motif (see [Cis-acting sequence involved in specific incorporation of selenocysteine at UGA codons of selenoprotein mRNAs](#)). The two common 3' UTR polymorphisms in the type 1 deiodinase gene (*DIO1*; MIM# 147982; Peeters et al. 2003, 2005) are not predicted to affect the SECIS motif and will thus be analysed together with the other group 4 variants. In addition, the polymorphism in the *GPX4* gene has not been functionally characterised and was therefore also included in group 4.

SEPI5

As illustrated in Fig. 5, the G/A1125 polymorphism does not affect the main stem-loop structure of the SECIS motif. However, it does serve to alter the sub-stem-loop structure attached to the apical loop. [Consistent with this, *mfold* analysis suggested that this polymorphic variant gives rise to a pattern I secondary structural change (Supplementary Fig. S11)]. This change may indirectly modify the three-dimensional structure of the SECIS motif, resulting in reduced binding of *trans*-acting factors such as SBP2. Indeed, using a specialised reporter gene construct, it has been demonstrated that although the A1125-containing SECIS motif was approximately twice as efficient in stimulating the readthrough of the UGA codon, it was less responsive to added selenium in the culture medium than the G1125-containing SECIS motif (Kumaraswamy et al. 2000; Hu et al. 2001). Malignant mesothelioma cells with the A1125 genotype were also less responsive to the growth inhibitory and apoptotic effects of selenium than cells with the G1125 genotype (Apostolou et al. 2004).

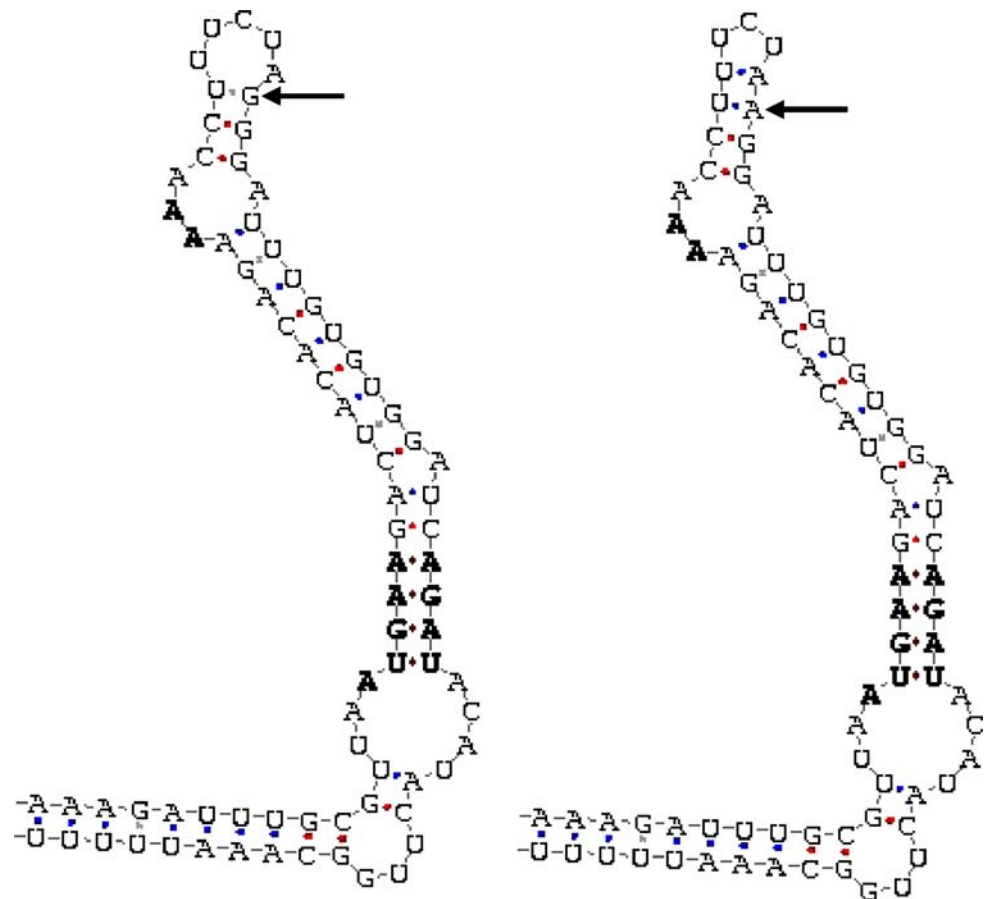
SEPN1

Very recently, a homozygous T>C mutation that occurred in the quartet 5'-UGAU-3':5'-AGAU-3' of the *SEPN1* gene has been detected in a patient with a mild form of rigid spine muscular dystrophy. This mutation was found to abolish SBP2 binding in vitro (Allamand et al. 2006) and results in a pattern II secondary structural change (Supplementary Fig. S12).

SLITRK1: a 3' UTR variant reported to occur within a miRNA target site

A G>A transition, in a predicted binding site for human miRNA hsa-miR-189 within the 3' UTR of the Slit and Trk-like 1 gene (*SLITRK1*; MIM# 609678), was identified in two apparently unrelated Tourette syndrome patients but was absent in 4,296 control chromosomes (Abelson et al. 2005). Experimental confirmation of the functional effect of this mutation came from the demonstration that, in the presence of hsa-miR-189, in vitro constructs bearing the 3' UTR mutation increased repression of a reporter gene by comparison with the wild-type. Interestingly, the G>A mutation results in a pattern I secondary structural change (Supplementary Fig. S13).

Fig. 5 Predicted secondary structures of the wild-type (*left panel*) and variant (*right panel*) RNA sequences in relation to the 3' UTR G>A polymorphism (indicated by *arrows*) in the *SEP15* gene using SECISearch 2.19. The well-conserved AA in the apical loop, the non-Watson-Crick base-pairs (the quartet; 5'-UGAA-3':5'-AGAU-3'), and an A in the internal loop are highlighted in *bold*. Note that in the secondary structures predicted by *mfold*, the non-Watson-Crick base-pairs (the quartet) are single stranded (Supplementary Fig. S11)



Correlating potential functionality with secondary structural changes

Comparison of the patterns of secondary structural changes observed with both the functional and non-functional USS variants (Table 4) strongly suggests that secondary structural change can be used as a reliable indicator of functionality. More importantly, this analysis sheds some light on several fundamental issues. Firstly, the frequently adopted rules to define *cis*-regulatory elements, based upon sequence comparisons between orthologous genes and/or in vitro functional analysis, appear to be inadequate to the task. For example, a variant that occurs within a well-conserved sequence tract that is double-stranded in the predicted RNA secondary structure may well have functional consequences. It would, however, be inappropriate immediately to term this sequence tract a *cis*-acting element because the variant could simply have induced a structural change in a nearby single-stranded, bona fide regulatory motif. Secondly, there may be a minimal length requirement for *cis*-acting elements. A detailed evaluation of the loops presumed to serve as binding sites for *trans*-acting factors reveals

that they usually comprise at least four nucleotides (Table 4).

Predicting functionality of the remaining USS variants using the newly established rules

The functionality or non-functionality of the group 3 and 4 variants can be reasonably well predicted by application of the rules listed in Table 4 and briefly described in Table 1. Here we shall confine our discussion to some novel findings and several interesting examples.

Identification of a new pattern of secondary structure change

Five variants (*GFPT2*, *IL12B*, and *RNASE3* in group 3; *CDKN2A* and *ZNF627* in group 4) were predicted to result in an altered orientation of the global stem-loop structure, exemplified by the *IL12B* A/C SNP (Fig. 6). This kind of radical secondary structural change (termed pattern III) could well be of functional significance.

Table 4 Characteristic hallmarks of different patterns of secondary structural change with respect to functionality

Prediction of functionality	Pattern of secondary structural changes	Changes in loop(s)	Changes in stem(s)
Non-functional	Pattern 0	No	Substitution of one nucleotide by another that is capable of pairing with the same opposing nucleotide in a helical portion
Functional	Pattern I	Size changed ^a	Length changed
Functional	Pattern II	The variant only changes the number of motif-defining nucleotides within a given loop ^b	No
Possibly functional	Pattern III	Altered orientation of the global stem-loop structure	Altered orientation of the global stem-loop structure
Unknown significance	Type 1	Yes, but change is more minor than for pattern I	Yes, but change is more minor than for pattern I
Unknown significance	Type 2	The involved loop comprises ≤ 3 nucleotides	No

Patterns 0–II were derived from functionally characterised USS variants, pattern III was derived mainly from group 3 USS variants, whereas type 1 and type 2 of unknown significance were virtually all derived from group 4 variants (refer to Table 1)

^aAt least one of the involved loops (either in the wild-type or variant) comprises at least four nucleotides

^bThe involved loop comprises at least four nucleotides

Further secondary structural patterns of unknown significance

Type 1

The secondary structural change predicted to be associated with the *AGTRI* A1166C variant (in group 4) involved the replacement of the paired A•U in the wild-type allele by a 1 × 1 internal loop in the variant allele (Supplementary Fig. S14). This secondary structural change as well as those of the *ASIP* g.8818A>G polymorphism (group 4; Supplementary Fig. S15) and the *DPYSL2* 2236T>C polymorphism (group 4; Supplementary Fig. S16) can in principle be assigned to pattern I. However, these changes are more minor than those of the pattern I functional variants and are thus termed patterns of unknown significance, type 1 (Table 4).

Type 2

Seven variants (*VEGF* 936C>T in group 3; *APC* A8822G, *ENDRA*, *KCNJ9* 8639A>G, *LEP*, *OLRI*, and *TCPI* in group 4) could in principle be assigned to pattern II. However, unlike the pattern II functional variants, all of which involve a loop (either hairpin or internal) comprising at least four nucleotides (Table 4), these seven variants involved an internal loop comprising only 1–3 nucleotides (e.g. Supplementary Fig. S17). These changes were thus termed patterns of unknown significance, type 2 (Table 4).

In the light of the above predictions, it would be interesting to investigate the functional consequences of these type 1 and 2 variants. Here it is perhaps worth reiterating that all these patterns of unknown significance were virtually identified in group 4 variants; the only one (i.e. *VEGF* 936C>T) from group 3 is in LD with a 1451C>T polymorphism that results in a pattern II change involving a bulge loop comprising five nucleotides (ACACC>ACACU).

Identification of potential novel *cis*-regulatory elements

In addition to the known AU-, AG-, C-rich and AC-rich motifs, some putatively novel USS motifs including AGCCUG(C/U)AG (*CDKN1A*) and a GC-rich sequence (*THPO*) were identified during the course of this analysis (Table 1).

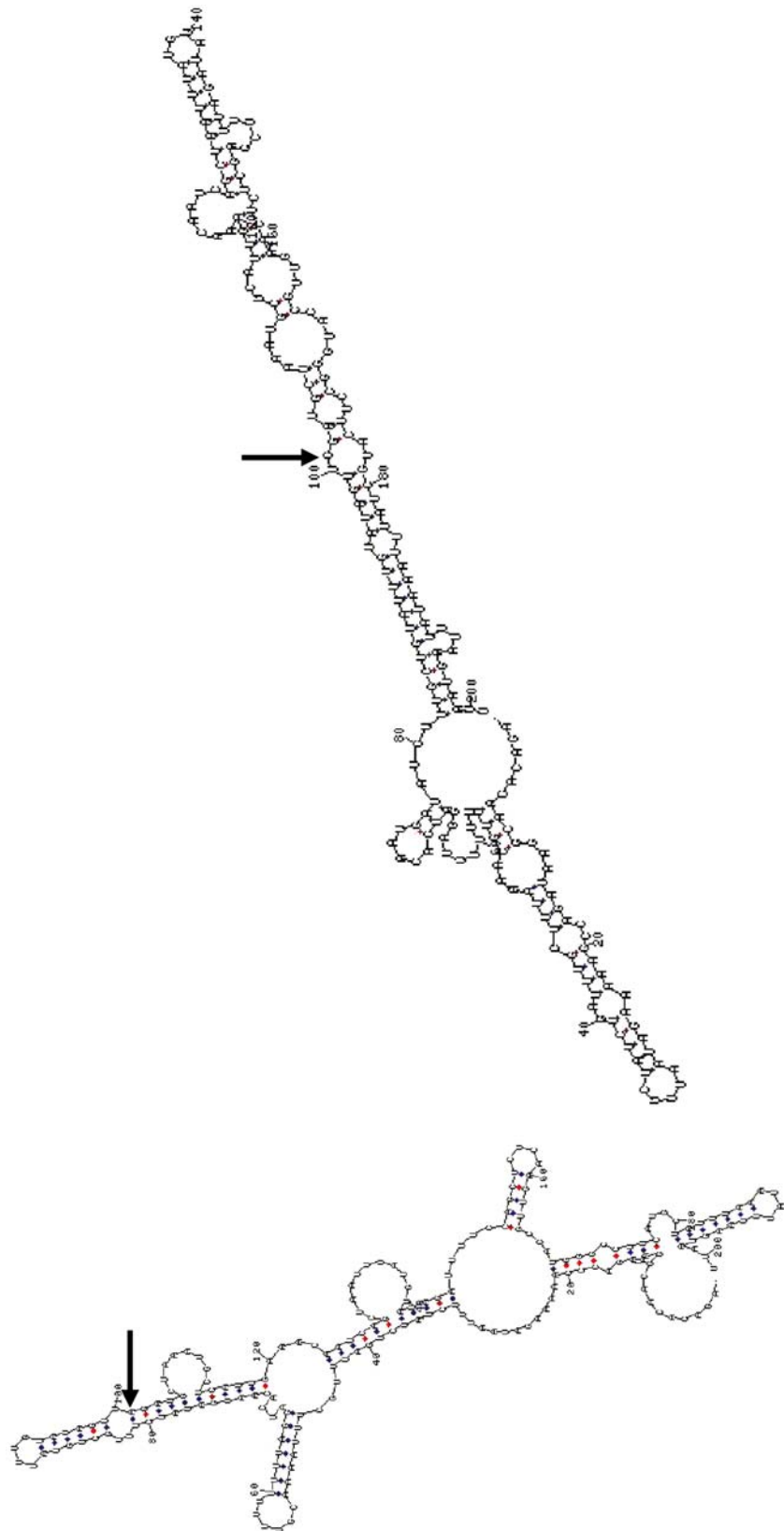
Several interesting examples

These examples in groups 3 and 4 are intended to stimulate discussion rather than to allow conclusions to be drawn.

CRP: *CRP* 3 (pattern I) vs. *CRP* 4 (pattern 0)

A haplotype containing the minor A allele of a G/A polymorphism (termed *CRP* 4) in the 3' UTR of the C-reactive protein (*CRP*) gene (MIM# 123260) has been

Fig. 6 Proposed secondary structures of the wild-type (*left panel*) and variant (*right panel*) RNA sequences in relation to the 3' UTR A>C polymorphism (position 101; indicated by *arrow*) in the *IL12B* gene. This kind of secondary structural change was termed pattern III



reported to be associated with both reduced basal CRP expression and the development of systematic lupus erythematosus (Russell et al. 2004). However, secondary structure prediction (pattern 0; Supplementary Fig. S18) suggests that CRP 4 is unlikely to be of functional significance. In this regard, CRP4 is in tight LD with another polymorphism in the 3' UTR of the *CRP* gene, CRP3 (CRP3 is located 858 bp upstream CRP4; Russell et al. 2004). CRP3 is identical to the +1444C>T polymorphism reported to influence in vivo CRP levels in other two studies (Brull et al. 2003; Kovacs et al. 2005). Thus, it is CRP3 that is probably of functional significance since it results in a pattern II secondary structure change in which the single-stranded AAACGG sequence in an internal loop is altered to AAAUGG (Supplementary Fig. S19).

GPX4

A common T/C polymorphism (allele frequencies 0.54/0.46) has been found in the 3' UTR of the *GPX4* gene (Villette et al. 2002). That individuals with different genotypes exhibited significant differences in the levels of lymphocyte 5-lipoxygenase suggests that this SNP may be of functional significance. As suggested by Supplementary Figs. S20 and S21, this common SNP probably affects the binding of *trans*-acting factors to the SECIS motif by modifying the three-dimensional structure of the *GPX4* mRNA.

IGF2: does the *ApaI* polymorphism affect the specific cleavage of *IGF2* mRNA?

As discussed earlier, specific cleavage of human *IGF2* mRNA is largely determined by two flanking stem structures, one of which involves a long-range RNA interaction between two widely separated (~2 kb) elements, I and II. It is perhaps pertinent to mention that the formation of this long-range stem structure is highly unusual and is facilitated firstly by the presence of a large tract (700–800 bp) of simple CA repeats and secondly by an extremely C-rich element II complementary to element I which is extremely G-rich (Supplementary Table S3).

An *ApaI* polymorphic site, initially detected at the *IGF2* cDNA level by Xiang et al. (1988), was deduced to be a G>A substitution (Tadokoro et al. 1991) at position 820 (in accordance with GenBank accession number X07868.1; Supplementary Table S3) within the USS. This polymorphism has been reported to be associated with body weight (O'Dell et al. 1997; Gaunt et al. 2001; Gu et al. 2002; Roth et al. 2002; Schragger

et al. 2004; Gomes et al. 2005) and to affect IGFII levels in vivo (O'Dell et al. 1997). Although the nucleotide at position 820 (located in the middle of a ~450 base tract between the element II and the simple CA repeats) is not involved in the formation of the two flanking stem structures, the G/A polymorphism does change the local secondary structure of the *IGF2* mRNA significantly (Pattern I; Supplementary Fig. S22). It is possible that this secondary structural change somehow impacts on the accessibility of the special ECS within the *IGF2* 3' UTR by modifying either or both of the two flanking stem structures.

IL12B: although the initial disease association of a 3' UTR A/C SNP may be spurious, this variant may nevertheless still act as a disease modifier

The initial report of the positive association of an A/C SNP in the 3' UTR of the *IL12B* gene (MIM# 161561) with type 1 diabetes (Morahan et al. 2001) is now generally regarded as a false positive (e.g. Dahlman et al. 2002). This notwithstanding, there are consistent and independent findings to suggest that this SNP does indeed have functional consequences; whilst allele C correlated with significantly increased secretion of IL12 protein after stimulation of human peripheral blood mononuclear cells (PBMCs) with *Staphylococcus aureus* plus recombinant human interferon- γ (Seegers et al. 2002), lipopolysaccharide (Yilmaz et al. 2005), or purified protein derivative (Yilmaz et al. 2005), the A allele correlated with increased IL12 p40 production in C3-binding glycoprotein-stimulated human PBMCs (Stanilova and Miteva 2005). However, given the fact that the A/C SNP is a very common polymorphism (minor C allele frequency of 20% in controls; Morahan et al. 2001), its clinical consequences, if there are any, are likely to be fairly minor. Interestingly, the authors who reported the initial positive association have recently reported that the C allele is associated with late onset type 1 diabetes: the C allele was found to be more common in patients diagnosed after the age of 16 than in controls or in patients diagnosed before the age of 16 (Windsor et al. 2004). Although this observation should still be regarded as preliminary until independently replicated, it may well provide a plausible molecular mechanism for mediating disease severity (as distinct from mediating disease susceptibility). In other words, it is possible that the *IL12B* A/C SNP may act as a disease modifier.

This postulate has been supported by several recent studies. Given the central role of T cells in viral control

and clearance in general (reviewed by Bowen and Walker 2005; Klenerman and Hill 2005) and the pivotal role of IL12 in the generation of the Th1 response in particular, three groups have postulated that the *IL12B* A/C SNP could affect spontaneous and treatment-induced recovery from hepatitis V virus (HCV)-infection. Their results [the A/A genotype was found to be associated with persistent infection (Yin et al. 2004; Houldsworth et al. 2005) whereas the C allele was associated with a more efficient response to antiviral combination therapy as a consequence of a reduced relapse gap (Mueller et al. 2004)] suggests that the C allele may act as a protective factor with respect to the outcome of HCV infection. However, the A/C SNP does not appear to be a susceptibility factor for HCV infection since its distribution did not significantly differ between patients and healthy controls (Mueller et al. 2004).

Taken together, the above observations strongly suggest that the *IL12B* A/C SNP is potentially of functional significance. This polymorphism causes a pattern III secondary structure change (Fig. 6).

OLRI: a 3' UTR variant that is in linkage disequilibrium with splicing variants

The *OLRI* gene (MIM# 602601) contains six exons and encodes oxidized low-density lipoprotein receptor 1. Whether a C>T SNP—located 188 bp 3' to the translational termination codon in the *OLRI* gene—is associated with atherosclerosis (Trabetti et al. 2006 and references therein) and/or Alzheimer's disease (Shi et al. 2006 and references therein) remains controversial. This SNP is, however, in complete LD with another five intronic SNPs: IVS4+27G>C, IVS4-73C>T, IVS4-14A>G, IVS5-70A>G, and IVS5-27G>T (Mango et al. 2003). Using quantitative real-time PCR and a minigene approach, Mango et al. (2005) demonstrated that the intronic variants regulate the expression of an exon 5-lacking isoform (termed *LOXIN*) of the *OLRI* gene, with the 'non-risk haplotype' being associated with increased expression of *LOXIN* at both the mRNA and protein levels. Further in vivo analysis suggested that increased expression of *LOXIN* protects cells from OLR1-induced apoptosis (Mango et al. 2005). Thus, the 'risk haplotype', irrespective of its potential medical significance, is biologically important owing to the presence of the multiple intronic variants. Interestingly, the *OLRI* 3' UTR C>T SNP results in a pattern of unknown significance, type 2 (Supplementary Fig. S17).

PHB: a controversial cancer risk-associated polymorphism exhibits a pattern 0 secondary structural change

A single C>T transition at position 729 of the *PHB* 3'-UTR (nomenclature in accordance with Fig. 5 in Jupe et al. 1996) has been reported to be associated with an increased risk of breast cancer in North American women, especially in those aged ≤ 50 years who already had a first-degree relative with the disease (Jupe et al. 2001). In support of this association, stable clones of the breast cancer cell line MCF7 transfected with plasmids containing the *PHB* wild-type 3'-UTR (i.e. the C allele) manifested significant suppression of growth in cell proliferation assays, inhibition of colony formation in soft agar assays, and suppression of xenograft tumour growth when implanted on nude mice, as compared with those clones transfected with either empty vectors or plasmids containing the *PHB* variant 3'-UTR (i.e. the T allele; Manjeshwar et al. 2003). Based upon these and their own previous results, Manjeshwar et al. (2003) stated that "our studies showing the cell cycle inhibitory activity of the prohibitin 3' UTR in a panel of immortalized cell lines certainly suggests that its loss of function may play a role in a number of cancers". However, not only was the association of the T allele with an increased risk of breast cancer not confirmed in two subsequent studies (Spurdle et al. 2002; Campbell et al. 2003), but also an association of this allele with risk of ovarian cancer has been excluded (Spurdle et al. 2003). A recent study from the group that reported the initial disease association and who performed the relevant functional evaluation of the C>T variant has, however, suggested that the *PHB* and other genes might interact to influence breast cancer risk in a manner not entirely predictable from single gene effects (Aston et al. 2005). Finally, it is pertinent to note that the C>T polymorphism results in a pattern 0 secondary structural change (Supplementary Fig. S23).

Isolated examples

ADRA2C polymorphisms: exemplifying the diverse intragenic variability with haplotype-specific functional effects

A 12-bp in-frame deletion/insertion polymorphism of nucleotides 964–975 in the presynaptic $\alpha 2C$ adrenergic receptor gene (*ADRA2C*; MIM# 104250), leading to the gain or loss of amino acids 322–325, constitutes a

risk factor for heart failure (Small et al. 2004 and references therein). This polymorphism was, however, partitioned between nine haplotypes that displayed differential mRNA and protein expression profiles in a whole-gene transfection assay. Obviously, in a given del322-325-containing haplotype, other polymorphic variants including a 21-bp insertion/deletion polymorphism in the 3'-UTR may serve to amplify, attenuate, or even dominate the phenotypic expression previously attributed solely to the del322-325 variant (Small et al. 2004).

GPR44: haplotypes containing two closely linked common 3' UTR polymorphisms

Two common polymorphisms, separated by only six nucleotides, occur in the 3' UTR of the *GPR44* gene (MIM# 604837) (Huang et al. 2004). Transcription pulsing experiments have shown that the 1544G-1651G haplotype confers a higher level of mRNA stability than the 1544C-1651A haplotype. However, as yet, it is unclear if this effect is conferred by one single variant or the two variants acting in concert.

TNFRSF1B: haplotypes containing different combinations of three SNPs within a 28 bp 3' UTR sequence tract

The *TNFRSF1B* (MIM# 191191) gene encodes the human TNF α receptor. Different combinations of three SNPs within a 28 bp sequence tract in the 3' UTR of *TNFRSF1B* constitute five different haplotypes. These haplotypes display different mRNA stabilities in transient expression assays (Puga et al. 2005).

VDR: a functional 3' UTR haplotype containing multiple variants

Polymorphisms of the vitamin D receptor gene (VDR; MIM# 601769) have been reported to be associated with increased fracture risk (Fang et al. 2005). Multiple polymorphic variants occur in the 3' UTR and manifest as several haplotypes. When evaluated using reporter gene constructs containing the complete 3.2 kb 3' UTR of VDR, the decay rate of haplotype 1-related mRNA was found to be 30% greater than that for haplotype 2-related mRNA (Fang et al. 2005). The individual contribution of each sequence variant is, however, unknown.

Concluding remarks

The idea that an RNA regulatory motif relies on a combination of primary and secondary structure is not

new. Indeed, more and more studies have incorporated secondary structure prediction in order to identify regulatory elements and/or to understand the nature of disease-associated regulatory variants. However, the systematic evaluation of known naturally occurring functional and non-functional USS variants has never been attempted before. In the present work, we have not only established a reliable and objective means to perform secondary structure prediction but, somewhat unexpectedly, we obtained consistent patterns of secondary structural change that appear to allow the discrimination of functional USS variants from their non-functional counterparts. The resulting rules were then employed to predict the potential functionality of the other collated USS variants. This notwithstanding, the validity and reliability of the predicted secondary structural patterns 0–III will need to be validated prospectively as new 3' UTR variants are identified and functionally analysed. Moreover, the predicted types 1 and 2 of unknown significance warrant further exploration. Nevertheless, had this type of analysis been performed by Winkler et al. (1998), they might well have been more cautious with respect to their interpretation regarding the biomedical consequences of the *CXCL12* SDF-G801A polymorphism.

Acknowledgements JMC is a visiting Professor of Genetics supported by the Ministère de la Jeunesse, de l'Éducation Nationale et de la Recherche, France. This work was supported by the INSERM (Institut National de la Santé et de la Recherche Médicale), France. The authors are most grateful to Peter Stenson for provision of HGMD data.

References

- Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, Mathews CA, Pauls DL, Rasin MR, Gunel M, Davis NR, Ercan-Sencicek AG, Guez DH, Spertus JA, Leckman JF, Dure LS IV, Kurlan R, Singer HS, Gilbert DL, Farhi A, Louvi A, Lifton RP, Sestan N, State MW (2005) Sequence variants in *SLITRK1* are associated with Tourette's syndrome. *Science* 310:317–320
- Adereth Y, Dammai V, Kose N, Li R, Hsu T (2005) RNA-dependent integrin alpha3 protein localization regulated by the Muscleblind-like protein MLP1. *Nat Cell Biol* 7:1140–1147
- Ainger K, Avossa D, Diana AS, Barry C, Barbarese E, Carson JH (1997) Transport and localization elements in myelin basic protein mRNA. *J Cell Biol* 138:1077–1087
- Aissouni Y, Perez C, Calmels B, Benech PD (2002) The cleavage/polyadenylation activity triggered by a U-rich motif sequence is differently required depending on the poly(A) site location at either the first or last 3'-terminal exon of the 2'-5' oligo(A) synthetase gene. *J Biol Chem* 277:35808–35814
- Allamand V, Richard P, Lescure A, Ledeuil C, Desjardin D, Petit N, Gartioux C, Ferreira A, Krol A, Pellegrini N, Urtizberea JA, Guichenev P (2006) A single homozygous

- point mutation in a 3' untranslated region motif of selenoprotein N mRNA causes SEPN1-related myopathy. *EMBO Rep* 7:450–454
- Al-Maghrebi M, Brule H, Padkina M, Allen C, Holmes WM, Zehner ZE (2002) The 3' untranslated region of human vimentin mRNA interacts with protein complexes containing eEF-1gamma and HAX-1. *Nucleic Acids Res* 30:5017–5028
- Apostolou S, Klein JO, Mitsuchi Y, Shetler JN, Poulikakos PI, Jhanwar SC, Kruger WD, Testa JR (2004) Growth inhibition and induction of apoptosis in mesothelioma cells by selenium and dependence on selenoprotein *SEPI5* genotype. *Oncogene* 23:5032–5240
- Aranda-Abreu GE, Hernandez ME, Soto A, Manzo J (2005) Possible *cis*-acting signal that could be involved in the localization of different mRNAs in neuronal axons. *Theor Biol Med Model* 2:33
- Arya SK, Ginsberg CC, Davis-Warren A, D'Costa J (1999) *In vitro* phenotype of *SDF1* gene mutant that delays the onset of human immunodeficiency virus disease *in vivo*. *J Hum Virol* 2:133–138
- Aston CE, Ralph DA, Lalo DP, Manjeshwar S, Gramling BA, DeFreese DC, West AD, Branam DE, Thompson LF, Craft MA, Mitchell DS, Shimasaki CD, Mulvihill JJ, Jupe ER (2005) Oligogenic combinations associated with breast cancer risk in women under 53 years of age. *Hum Genet* 116:208–221
- Audrezet MP, Chen JM, Le Marechal C, Ruzsniowski P, Robaszkiwicz M, Raguene O, Quere I, Scotet V, Ferec C (2002) Determination of the relative contribution of three genes - the cystic fibrosis transmembrane conductance regulator gene, the cationic trypsinogen gene, and the pancreatic secretory trypsin inhibitor gene-to the etiology of idiopathic chronic pancreatitis. *Eur J Hum Genet* 10:100–106
- Baker M, Gaukrodger N, Mayosi BM, Imrie H, Farrall M, Watkins H, Connell JM, Avery PJ, Keavney B (2005) Association between common polymorphisms of the proopiomelanocortin gene and body fat distribution: a family study. *Diabetes* 54:2492–2496
- Bakheet T, Frevel M, Williams BR, Greer W, Khabar KS (2001) ARED: human AU-rich element-containing mRNA database reveals an unexpectedly diverse functional repertoire of encoded proteins. *Nucleic Acids Res* 29:246–254
- Barreau C, Paillard L, Osborne HB (2006) AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res* 33:7138–7150
- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116:281–297
- Basak AN, Ozer A, Kirdar B, Akar N (1993) A novel 13 bp deletion in the 3' UTR of the beta-globin gene causes beta-thalassemia in a Turkish patient. *Hemoglobin* 17:551–555
- Becker KF, Reich U, Schott C, Hofler H (1995) Single nucleotide polymorphisms in the human E-cadherin gene. *Hum Genet* 96:739–740
- Beffagna G, Occhi G, Nava A, Vitiello L, Ditadi A, Basso C, Bauce B, Carraro G, Thiene G, Towbin JA, Danieli GA, Rampazzo A (2005) Regulatory mutations in transforming growth factor-beta3 gene cause arrhythmogenic right ventricular cardiomyopathy type 1. *Cardiovasc Res* 65:366–373
- Benjafield AV, Katyk K, Morris BJ (2003) Association of *ED-NRA*, but not *WNK4* or *FKBP1B*, polymorphisms with essential hypertension. *Clin Genet* 64:433–438
- Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E, Sharon E, Spector Y, Bentwich Z (2005) Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet* 37:766–770
- Berger A, Stierkorb E, Nickenig G (2005) The role of the AUUUUA hexamer for the posttranscriptional regulation of the AT1 receptor mRNA stability. *Biochem Biophys Res Commun* 330:805–812
- Bertram L, Parkinson M, McQueen MB, Mullin K, Hsiao M, Menon R, Moscarillo TJ, Blacker D, Tanzi RE (2005) Further evidence for LBP-1c/CP2/LSF association in Alzheimer's disease families. *J Med Genet* 42:857–862
- Berry MJ (2005) Insights into the hierarchy of selenium incorporation. *Nat Genet* 37:1162–1163
- Bilenoglu O, Basak AN, Russell JE (2002) A 3' UTR mutation affects beta-globin expression without altering the stability of its fully processed mRNA. *Br J Haematol* 119:1106–1114
- Binder R, Horowitz JA, Basilion JP, Koeller DM, Klausner RD, Harford JB (1994) Evidence that the pathway of transferrin receptor mRNA degradation involves an endonucleolytic cleavage within the 3' UTR and does not involve poly(A) tail shortening. *EMBO J* 13:1969–1980
- Bonilla C, Boxill LA, Donald SA, Williams T, Sylvester N, Parra EJ, Dios S, Norton HL, Shriver MD, Kittles RA (2005) The 8818G allele of the agouti signaling protein (*ASIP*) gene is ancestral and is associated with darker skin color in African Americans. *Hum Genet* 116:402–406
- Bonnardeaux A, Davies E, Jeunemaitre X, Fery I, Charru A, Clauser E, Tiret L, Cambien F, Corvol P, Soubrier F (1994) Angiotensin II type 1 receptor gene polymorphisms in human essential hypertension. *Hypertension* 24:63–69
- Bowen DG, Walker CM (2005) Adaptive immune responses in acute and chronic hepatitis C virus infection. *Nature* 436:946–952
- Brackenridge S, Proudfoot NJ (2000) Recruitment of a basal polyadenylation factor by the upstream sequence element of the human lamin B2 polyadenylation signal. *Mol Cell Biol* 20:2660–2669
- Brackenridge S, Ashe HL, Giacca M, Proudfoot NJ (1997) Transcription and polyadenylation in a short human intergenic region. *Nucleic Acids Res* 25:2326–2336
- Branson R, Potoczna N, Kral JG, Lentjes KU, Hoehe MR, Horber FF (2003) Binge eating as a major phenotype of melanocortin 4 receptor gene mutations. *New Engl J Med* 348:1096–1103
- Brown CY, Lagnado CA, Goodall GJ (1996) A cytokine mRNA-destabilizing element that is structurally and functionally distinct from A+U-rich elements. *Proc Natl Acad Sci USA* 93:13721–13725
- Brull DJ, Serrano N, Zito F, Jones L, Montgomery HE, Rumley A, Sharma P, Lowe GD, World MJ, Humphries SE, Hingorani AD (2003) Human *CRP* gene polymorphism influences CRP levels: implications for the prediction and pathogenesis of coronary heart disease. *Arterioscler Thromb Vasc Biol* 23:2063–2069
- Campa D, Zienolddiny S, Maggini V, Skaug V, Haugen A, Canzian F (2004) Association of a common polymorphism in the cyclooxygenase 2 gene with risk of non-small cell lung cancer. *Carcinogenesis* 25:229–235
- Campbell IG, Allen J, Eccles DM (2003) Prohibitin 3' untranslated region polymorphism and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 12:1273–1274
- Cappellini MD, Martinez di Montemuros F, Tavazzi D, Fargion S, Pizzuti A, Comino A, Cainelli T, Fiorelli G (2001) Seven novel point mutations in the uroporphyrinogen decarboxylase (*UROD*) gene in patients with familial porphyria cutanea tarda (f-PCT). *Hum Mutat* 17:350
- Caput D, Beutler B, Hartog K, Thayer R, Brown-Shimer S, Cerami A (1986) Identification of a common nucleotide

- sequence in the 3'-untranslated region of mRNA molecules specifying inflammatory mediators. *Proc Natl Acad Sci USA* 83:1670–1674
- Carlton VE, Hu X, Chokkalingam AP, Schrodri SJ, Brandon R, Alexander HC, Chang M, Catanese JJ, Leong DU, Ardlie KG, Kastner DL, Seldin MF, Criswell LA, Gregersen PK, Beasley E, Thomson G, Amos CI, Begovich AB (2005) *PTPN22* genetic variation: evidence for multiple variants associated with rheumatoid arthritis. *Am J Hum Genet* 77:567–581
- Casey JL, Hentze MW, Koeller DM, Caughman SW, Rouault TA, Klausner RD, Harford JB (1988) Iron-responsive elements: regulatory RNA sequences that control mRNA levels and translation. *Science* 240:924–928
- Casey JL, Koeller DM, Ramin VC, Klausner RD, Harford JB (1989) Iron regulation of transferrin receptor mRNA levels requires iron-responsive elements and a rapid turnover determinant in the 3' untranslated region of the mRNA. *EMBO J* 8:3693–3699
- Chabanon H, Mickleburgh I, Hesketh J (2004) Zipcodes and postage stamps: mRNA localisation signals and their *trans*-acting binding proteins. *Brief Funct Genomic Proteomic* 3:240–256
- Chabanon H, Mickleburgh I, Burtle B, Pedder C, Hesketh J (2005) An AU-rich stem-loop structure is a critical feature of the perinuclear localization signal of *c-myc* mRNA. *Biochem J* 392:475–483
- Charmandari E, Chrousos GP, Ichijo T, Bhattacharyya N, Vottero A, Souvatzoglou E, Kino T (2005) The human glucocorticoid receptor (hGR) beta isoform suppresses the transcriptional activity of hGRalpha by interfering with formation of active coactivator complexes. *Mol Endocrinol* 19:52–64
- Chen CY, Shyu AB (1995) AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem Sci* 20:465–470
- Chen CY, Gherzi R, Ong SE, Chan EL, Raijmakers R, Puijn GJ, Stoecklin G, Moroni C, Mann M, Karin M (2001) AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell* 107:451–464
- Chen JM, Férec C, Cooper DN (2006) A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes I: General principles and overview. *Hum Genet* [Epub ahead of print]
- Christian K, Lang M, Maurel P, Raffalli-Mathieu F (2004) Interaction of heterogeneous nuclear ribonucleoprotein A1 with cytochrome P450 2A6 mRNA: implications for post-transcriptional regulation of the *CYP2A6* gene. *Mol Pharmacol* 65:1405–1414
- Comings DE, Wu S, Rostamkhani M, McGue M, Iacono WG, MacMurray JP (2002) Association of the muscarinic cholinergic 2 receptor (*CHRM2*) gene with major depression in women. *Am J Med Genet* 114:527–529
- Cox DG, Pontes C, Guino E, Navarro M, Osorio A, Canzian F, Moreno V; Bellvitge Colorectal Cancer Study Group (2004) Polymorphisms in prostaglandin synthase 2/cyclooxygenase 2 (*PTGS2/COX2*) and risk of colorectal cancer. *Br J Cancer* 91:339–343
- Culjkovic B, Topisirovic I, Skrabanek L, Ruiz-Gutierrez M, Borden KL (2005) eIF4E promotes nuclear export of cyclin D1 mRNAs via an element in the 3'UTR. *J Cell Biol* 169:245–256
- Dahlman I, Eaves IA, Kosoy R, Morrison VA, Heward J, Gough SC, Allahabadia A, Franklyn JA, Tuomilehto J, Tuomilehto-Wolf E, Cucca F, Guja C, Ionescu-Tirgoviste C, Stevens H, Carr P, Nutland S, McKinney P, Shield JP, Wang W, Cordell HJ, Walker N, Todd JA, Concannon P (2002) Parameters for reliable results in genetic association studies in common disease. *Nat Genet* 30:149–150
- Dahlman I, Wahrenberg H, Persson L, Arner P (2004) No association of reported functional protein tyrosine phosphatase 1B 3' UTR gene polymorphism with features of the metabolic syndrome in a Swedish population. *J Intern Med* 255:694–695
- Dammerman M, Sandkuijl LA, Halaas JL, Chung W, Breslow JL (1993) An apolipoprotein CIII haplotype protective against hypertriglyceridemia is specified by promoter and 3' untranslated region polymorphisms. *Proc Natl Acad Sci USA* 90:4562–4566
- Deinum J, van Gool JM, Kofflard MJ, ten Cate FJ, Danser AH (2001) Angiotensin II type 2 receptors and cardiac hypertrophy in women with hypertrophic cardiomyopathy. *Hypertension* 38:1278–1281
- Derijk RH, Schaaf MJ, Turner G, Datson NA, Vreugdenhil E, Cidlowski J, de Kloet ER, Emery P, Sternberg EM, Detera-Wadleigh SD (2001) A human glucocorticoid receptor gene variant that increases the stability of the glucocorticoid receptor beta-isoform mRNA is associated with rheumatoid arthritis. *J Rheumatol* 28:2383–2388
- van Dijk EL, Sussenbach JS, Holthuizen PE (1998) Identification of RNA sequences and structures involved in site-specific cleavage of *IGF-II* mRNAs. *RNA* 4:1623–1635
- van Dijk EL, Sussenbach JS, Holthuizen PE (2000) Distinct RNA structural domains cooperate to maintain a specific cleavage site in the 3'-UTR of *IGF-II* mRNAs. *J Mol Biol* 300:449–467
- Di Paola R, Frittitta L, Miscio G, Bozzali M, Baratta R, Centra M, Spampinato D, Santagati MG, Ercolino T, Cisternino C, Soccio T, Mastroianno S, Tassi V, Almgren P, Pizzuti A, Vigneri R, Trischitta V (2002) A variation in 3' UTR of *hPTPIB* increases specific gene expression and associates with insulin resistance. *Am J Hum Genet* 70:806–812
- Dixon DA, Balch GC, Kedersha N, Anderson P, Zimmerman GA, Beauchamp RD, Prescott SM (2003) Regulation of cyclooxygenase-2 expression by the translational silencer TIA-1. *J Exp Med* 198:475–481
- Doi K, Noiri E, Nakao A, Fujita T, Kobayashi S, Tokunaga K (2006) Functional polymorphisms in the vascular endothelial growth factor gene are associated with development of end-stage renal disease in males. *J Am Soc Nephrol* 17:823–830
- Echwald SM, Bach H, Vestergaard H, Richelsen B, Kristensen K, Drivsholm T, Borch-Johnsen K, Hansen T, Pedersen O (2002) A P387L variant in protein tyrosine phosphatase-1B (*PTP-1B*) is associated with type 2 diabetes and impaired serine phosphorylation of PTP-1B *in vitro*. *Diabetes* 51:1–6
- Espel E (2005) The role of the AU-rich elements of mRNAs in controlling translation. *Semin Cell Dev Biol* 16:59–67
- Fang Y, van Meurs JB, d'Alesio A, Jhamai M, Zhao H, Rivadeneira F, Hofman A, van Leeuwen JP, Jehan F, Pols HA, Uitterlinden AG (2005) Promoter and 3'-untranslated-region haplotypes in the vitamin D receptor gene predispose to osteoporotic fracture: the Rotterdam study. *Am J Hum Genet* 77:807–823
- Fialcowitz EJ, Brewer BY, Keenan BP, Wilson GM (2005) A hairpin-like structure within an AU-rich mRNA-stabilizing element regulates *trans*-factor binding selectivity and mRNA decay kinetics. *J Biol Chem* 280:22406–22417
- Fitze G, Schierz M, Kuhlisch E, Schreiber M, Ziegler A, Roesner D, Schackert HK (2003) Novel intronic polymorphisms in the *RET* proto-oncogene and their association with Hirschsprung disease. *Hum Mutat* 22:177

- Florez JC, Agapakis CM, Burt NP, Sun M, Almgren P, Rastam L, Tuomi T, Gaudet D, Hudson TJ, Daly MJ, Ardlie KG, Hirschhorn JN, Groop L, Altshuler D (2005) Association testing of the protein tyrosine phosphatase 1B gene (*PTPNI*) with type 2 diabetes in 7,883 People. *Diabetes* 54:1884–1891
- Fritz DT, Jiang S, Xu J, Rogers MB (2006) A polymorphism in a conserved post-transcriptional regulatory motif alters BMP2 RNA protein interactions. *Mol Endocrinol* [Epub ahead of print]
- Gardier S, Vincent M, Lantelme P, Rial MO, Bricca G, Milon H (2004) A1166C polymorphism of angiotensin II type 1 receptor, blood pressure and arterial stiffness in hypertension. *J Hypertens* 22:2135–2142
- Garner C, Best S, Menzel S, Rooks H, Spector TD, Thein SL (2005) Two candidate genes for low platelet count identified in an Asian Indian kindred by genome-wide linkage analysis: glycoprotein IX and thrombopoietin. *Eur J Hum Genet* 14:101–108
- Gaukrodger N, Mayosi BM, Imrie H, Avery P, Baker M, Connell JM, Watkins H, Farrall M, Keavney B (2005) A rare variant of the leptin gene has large effects on blood pressure and carotid intima-medial thickness: a study of 1428 individuals in 248 families. *J Med Genet* 42:474–478
- Gaunt TR, Cooper JA, Miller GJ, Day IN, O'Dell SD (2001) Positive associations between single nucleotide polymorphisms in the *IGF2* gene region and body mass index in adult males. *Hum Mol Genet* 10:1491–1501
- Gladyshev VN, Jeang KT, Wootton JC, Hatfield DL (1998) A new human selenium-containing protein. Purification, characterization, and cDNA sequence. *J Biol Chem* 273:8910–8915
- Glisovic T, Ben-David Y, Lang MA, Raffalli-Mathieu F (2003) Interplay between hnRNP A1 and a *cis*-acting element in the 3' UTR of *CYP2A5* mRNA is central for high expression of the gene. *FEBS Lett* 535:147–152
- Gomes MV, Soares MR, Pasqualim-Neto A, Marcondes CR, Lobo RB, Ramos ES (2005) Association between birth weight, body mass index and *IGF2/ApaI* polymorphism. *Growth Horm IGF Res* 15:360–362
- Graziano F, Kawakami K, Watanabe G, Ruzzo A, Humar B, Santini D, Catalano V, Ficarelli R, Merriman T, Panunzi S, Testa E, Cascinu S, Bearzi I, Tonini G, Magnani M (2004) Association of thymidylate synthase polymorphisms with gastric cancer susceptibility. *Int J Cancer* 112:1010–1014
- Gu D, O'Dell SD, Chen XH, Miller GJ, Day IN (2002) Evidence of multiple causal sites affecting weight in the *IGF2-INS-TH* region of human chromosome 11. *Hum Genet* 110:173–181
- Gu J, Liang D, Wang Y, Lu C, Wu X (2005) Effects of *N*-acetyl transferase 1 and 2 polymorphisms on bladder cancer risk in Caucasians. *Mutat Res* 581:97–104
- Hall-Pogar T, Zhang H, Tian B, Lutz CS (2005) Alternative polyadenylation of cyclooxygenase-2. *Nucleic Acids Res* 33:2565–2579
- Han J, Brown T, Beutler B (1990) Endotoxin-responsive sequences control cachectin/tumor necrosis factor biosynthesis at the translational level. *J Exp Med* 171:465–475
- Han SW, Kim GW, Seo JS, Kim SJ, Sa KH, Park JY, Lee J, Kim SY, Goronzy JJ, Weyand CM, Kang YM (2004) *VEGF* gene polymorphisms and susceptibility to rheumatoid arthritis. *Rheumatology* 43:1173–1177
- Hashizume K, Mashima Y, Fumayama T, Ohtake Y, Kimura I, Yoshida K, Ishikawa K, Yasuda N, Fujimaki T, Asaoka R, Koga T, Kanamoto T, Fukuchi T, Miyaki K; Glaucoma Gene Research Group (2005) Genetic polymorphisms in the angiotensin II receptor gene and their association with open-angle glaucoma in a Japanese population. *Invest Ophthalmol Vis Sci* 46:1993–2001
- Heinimann K, Thompson A, Locher A, Furlanetto T, Bader E, Wolf A, Meier R, Walter K, Bauerfeind P, Marra G, Muller H, Foerzler D, Dobbie Z (2001) Nontruncating *APC* germ-line mutations and mismatch repair deficiency play a minor role in *APC* mutation-negative polyposis. *Cancer Res* 61:7616–7622
- Holcik M, Liebhaber SA (1997) Four highly stable eukaryotic mRNAs assemble 3' untranslated region RNA-protein complexes sharing *cis* and *trans* components. *Proc Natl Acad Sci USA* 94:2410–2414
- Hong LE, Wonodi I, Avila MT, Buchanan RW, McMahon RP, Mitchell BD, Stine OC, Carpenter WT Jr, Thaker GK (2005) Dihydropyrimidinase-related protein 2 (*DRP-2*) gene and association to deficit and nondeficit schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* 136:8–11
- Houldsworth A, Metzner M, Rossol S, Shaw S, Kaminski E, Demaine AG, Cramp ME (2005) Polymorphisms in the *IL-12B* gene and outcome of HCV infection. *J Interferon Cytokine Res* 25:271–276
- Hu YJ, Korotkov KV, Mehta R, Hatfield DL, Rotimi CN, Luke A, Prewitt TE, Cooper RS, Stock W, Vokes EE, Dolan ME, Gladyshev VN, Diamond AM (2001) Distribution and functional consequences of nucleotide polymorphisms in the 3'-untranslated region of the human *Sep15* gene. *Cancer Res* 61:2307–2310
- Hu J, Lutz CS, Wilusz J, Tian B (2005) Bioinformatic identification of candidate *cis*-regulatory elements involved in human mRNA polyadenylation. *RNA* 11:1485–1493
- Huang JL, Gao PS, Mathias RA, Yao TC, Chen LC, Kuo ML, Hsu SC, Plunkett B, Toggias A, Barnes KC, Stellato C, Beaty TH, Huang SK (2004) Sequence variants of the gene encoding chemoattractant receptor expressed on Th2 cells (*CRT2*) are associated with asthma and differentially influence mRNA stability. *Hum Mol Genet* 13:2691–2697
- Ioannidis JP, Rosenberg PS, Goedert JJ, Ashton LJ, Benfield TL, Buchbinder SP, Coutinho RA, Eugen-Olsen J, Gallart T, Katzenstein TL, Kostrikis LG, Kuipers H, Louie LG, Mallal SA, Margolick JB, Martinez OP, Meyer L, Michael NL, Operskalski E, Pantaleo G, Rizzardì GP, Schuitemaker H, Sheppard HW, Stewart GJ, Theodorou ID, Ullum H, Vicenzi E, Vlahov D, Wilkinson D, Workman C, Zagury JF, O'Brien TR; International Meta-Analysis of HIV Host Genetics (2001) Effects of *CCR5-Delta32*, *CCR2-64I*, and *SDF-1 3'A* alleles on HIV-1 disease progression: An international meta-analysis of individual-patient data. *Ann Intern Med* 135:782–795
- Jaeger S, Barends S, Giege R, Eriani G, Martin F (2005) Expression of metazoan replication-dependent histone genes. *Biochimie* 87:827–834
- Jankovic L, Dimovski AJ, Kollia P, Karageorga M, Loukopoulos D, Huisman TH (1991) A C → G mutation at nt position 6 3' to the terminating codon may be the cause of a silent beta-thalassemia. *Int J Hematol* 54:289–293
- Jeong BH, Kim NH, Choi EK, Lee C, Song YH, Kim JI, Carp RI, Kim YS (2005) Polymorphism at 3' UTR +28 of the prion-like protein gene is associated with sporadic Creutzfeldt-Jakob disease. *Eur J Hum Genet* 13:1094–1097
- Jones A, Dhamrait SS, Payne JR, Hawe E, Li P, Toor IS, Luong L, Wootton PT, Miller GJ, Humphries SE, Montgomery HE (2003) Genetic variants of angiotensin II receptors and cardiovascular risk in hypertension. *Hypertension* 42:500–506

- Jonsson UB, Bystrom J, Stalenheim G, Venge P (2006) A (G->C) transversion in the 3' UTR of the human *ECP* (eosinophil cationic protein) gene correlates to the cellular content of ECP. *J Leukoc Biol* 79:846–851
- Jupe ER, Liu XT, Kiehlbauch JL, McClung JK, Dell'Orco RT (1996) The 3' untranslated region of prohibitin and cellular immortalization. *Exp Cell Res* 224:128–135
- Jupe ER, Badgett AA, Neas BR, Craft MA, Mitchell DS, Resta R, Mulvihill JJ, Aston CE, Thompson LF (2001) Single nucleotide polymorphism in prohibitin 3' untranslated region and breast-cancer susceptibility. *Lancet* 357:1588–1589
- Kanetsky PA, Swoyer J, Panossian S, Holmes R, Guerry D, Rebbeck TR (2002) A polymorphism in the agouti signaling protein gene is associated with human pigmentation. *Am J Hum Genet* 70:770–775
- Katz LA, Schultz RE, Semina EV, Torfs CP, Krahn KN, Murray JC (2004) Mutations in *PITX2* may contribute to cases of omphalocele and VATER-like syndromes. *Am J Med Genet A* 130:277–283
- Kaufmann I, Martin G, Friedlein A, Langen H, Keller W (2004) Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO J* 23:616–626
- Khabar KS (2005) The AU-rich transcriptome: more than interferons and cytokines, and its role in disease. *J Interferon Cytokine Res* 25:1–10
- Khabar KS, Bakheet T, Williams BR (2005) AU-rich transient response transcripts in the human genome: expressed sequence tag clustering and gene discovery approach. *Genomics* 85:165–175
- Kiledjian M, Wang X, Liebhaber SA (1995) Identification of two KH domain proteins in the alpha-globin mRNP stability complex. *EMBO J* 14:4357–4364
- Kimura R, Nishioka T, Ishida T (2003) The *SDF1-G801A* polymorphism is not associated with *SDF1* gene expression in *Epstein-Barr* virus-transformed lymphoblastoid cells. *Genes Immun* 4:356–361
- Kimura R, Nishioka T, Soemantri A, Ishida T (2005) Allele-specific transcript quantification detects haplotypic variation in the levels of the *SDF-1* transcripts. *Hum Mol Genet* 14:1579–1585
- Kleiman FE, Ramirez AO, Dodelson de Kremer R, Gravel RA, Argarana CE (1998) A frequent TG deletion near the polyadenylation signal of the human *HEXB* gene: occurrence of an irregular DNA structure and conserved nucleotide sequence motif in the 3' untranslated region. *Hum Mutat* 12:320–329
- Klein B, Weirich G, Brauch H (2001) DHPLC-based germline mutation screening in the analysis of the *VHL* tumor suppressor gene: usefulness and limitations. *Hum Genet* 108:376–384
- Klenerman P, Hill A (2005) T cells and viral persistence: lessons from diverse infections. *Nat Immunol* 6:873–879
- Knowles JW, Erickson LM, Guy VK, Sigel CS, Wilder JC, Maeda N (2003) Common variations in noncoding regions of the human natriuretic peptide receptor A gene have quantitative effects. *Hum Genet* 112:62–70
- Koeller DM, Casey JL, Hentze MW, Gerhardt EM, Chan LN, Klausner RD, Harford JB (1989) A cytosolic protein binds to structural elements within the iron regulatory region of the transferrin receptor mRNA. *Proc Natl Acad Sci USA* 86:3574–3578
- Koeller DM, Horowitz JA, Casey JL, Klausner RD, Harford JB (1991) Translation and the stability of mRNAs encoding the transferrin receptor and *c-fos*. *Proc Natl Acad Sci USA* 88:7778–7782
- Korotkov KV, Novoselov SV, Hatfield DL, Gladyshev VN (2002) Mammalian selenoprotein in which selenocysteine (Sec) incorporation is supported by a new form of Sec insertion sequence element. *Mol Cell Biol* 22:1402–1411
- Kovacs A, Green F, Hansson LO, Lundman P, Samnegard A, Boquist S, Ericsson CG, Watkins H, Hamsten A, Tornvall P (2005) A novel common single nucleotide polymorphism in the promoter region of the C-reactive protein gene associated with the plasma concentration of C-reactive protein. *Atherosclerosis* 178:193–198
- Krippel P, Langsenlehner U, Renner W, Yazdani-Biuki B, Wolf G, Wascher TC, Paulweber B, Haas J, Samonigg H (2003) A common 936 C/T gene polymorphism of vascular endothelial growth factor is associated with decreased breast cancer risk. *Int J Cancer* 106:468–471
- Kruys V, Marinx O, Shaw G, Deschamps J, Huez G (1989) Translational blockade imposed by cytokine-derived U-rich sequences. *Science* 245:852–855
- Kryukov GV, Castellano S, Novoselov SV, Lobanov AV, Zehtab O, Guigo R, Gladyshev VN (2003) Characterization of mammalian selenoproteomes. *Science* 300:1439–1443
- Kumar R, Smeds J, Berggren P, Straume O, Rozell BL, Akhlen LA, Hemminki K (2001) A single nucleotide polymorphism in the 3' untranslated region of the *CDKN2A* gene is common in sporadic primary melanomas but mutations in the *CDKN2B*, *CDKN2C*, *CDK4* and *p53* genes are rare. *Int J Cancer* 95:388–393
- Kumaraswamy E, Malykh A, Korotkov KV, Kozyavkin S, Hu Y, Kwon SY, Moustafa ME, Carlson BA, Berry MJ, Lee BJ, Hatfield DL, Diamond AM, Gladyshev VN (2000) Structure-expression relationships of the 15-kDa selenoprotein gene. Possible role of the protein in cancer etiology. *J Biol Chem* 275:35540–35547
- Lambert JC, Goumidi L, Vrieze FW, Frigard B, Harris JM, Cummings A, Coates J, Pasquier F, Cottel D, Gaillac M, St Clair D, Mann DM, Hardy J, Lendon CL, Amouyel P, Chartier-Harlin MC (2000) The transcriptional factor LBP-1/CP2/LSF gene on chromosome 12 is a genetic determinant of Alzheimer's disease. *Hum Mol Genet* 9:2275–2280
- Legendre M, Gautheret D (2003) Sequence determinants in human polyadenylation site selection. *BMC Genomics* 4:7
- Leroy-Matheron C, Duchemin J, Levent M, Gouault-Heilmann M (1999) Genetic modulation of plasma protein S levels by two frequent dimorphisms in the *PROS1* gene. *Thromb Haemost* 82:1088–1092
- Liao Z, Liu H, Swisher SG, Wang L, Wu TT, Correa AM, Roth JA, Cox JD, Komaki R, Ajani JA, Wei Q (2006) Polymorphism at the 3'-UTR of the thymidylate synthase gene: A potential predictor for outcomes in Caucasian patients with esophageal adenocarcinoma treated with preoperative chemoradiation. *Int J Radiat Oncol Biol Phys* 64:700–708
- Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433:769–773
- Lin HJ, Tsai FJ, Hung P, Chen WC, Chen HY, Fan SS, Tsai SW (2006) Association of E-cadherin gene 3'-UTR C/T polymorphism with primary open angle glaucoma. *Ophthalmic Res* 38:44–48
- Lopez de Silanes I, Zhan M, Lal A, Yang X, Gorospe M (2004) Identification of a target RNA motif for RNA-binding protein HuR. *Proc Natl Acad Sci USA* 101:2987–2992
- Lopez de Silanes I, Galban S, Martindale JL, Yang X, Mazan-Mamczarz K, Indig FE, Falco G, Zhan M, Gorospe M (2005) Identification and functional outcome of mRNAs

- associated with RNA-binding protein TIA-1. *Mol Cell Biol* 25:9520–9531
- Lu JW, Gao CM, Wu JZ, Cao HX, Tajima K, Feng JF (2006) Polymorphism in the 3'-untranslated region of the thymidylate synthase gene and sensitivity of stomach cancer to fluoropyrimidine-based chemotherapy. *J Hum Genet* 51:155–160
- Luedeking-Zimmer E, DeKosky ST, Nebes R, Kamboh MI (2003) Association of the 3' UTR transcription factor LBP-1c/CP2/LSF polymorphism with late-onset Alzheimer's disease. *Am J Med Genet B Neuropsychiatr Genet* 117:114–117
- Mandola MV, Stoehlmacher J, Zhang W, Groshen S, Yu MC, Iqbal S, Lenz HJ, Ladner RD (2004) A 6 bp polymorphism in the thymidylate synthase gene causes message instability and is associated with decreased intratumoral TS mRNA levels. *Pharmacogenetics* 14:319–327
- Mango R, Clementi F, Borgiani P, Forleo GB, Federici M, Contino G, Giardina E, Garza L, Fahdi IE, Lauro R, Mehta JL, Novelli G, Romeo F (2003) Association of single nucleotide polymorphisms in the oxidised LDL receptor 1 (*OLRI*) gene in patients with acute myocardial infarction. *J Med Genet* 40:933–936
- Mango R, Biocca S, del Vecchio F, Clementi F, Sanguuolo F, Amati F, Filareto A, Grelli S, Spitalieri P, Filesi I, Favalli C, Lauro R, Mehta JL, Romeo F, Novelli G (2005) *In vivo* and *in vitro* studies support that a new splicing isoform of *OLRI* gene is protective against acute myocardial infarction. *Circ Res* 97:152–158
- Manjeshwar S, Branam DE, Lerner MR, Brackett DJ, Jupe ER (2003) Tumor suppression by the prohibitin gene 3'untranslated region RNA in human breast cancer. *Cancer Res* 63:5251–5256
- Maragoudaki E, Vrettou C, Kanavakis E, Traeger-Synodinos J, Metaxotou-Mavrommati A, Kattamis C (1998) Molecular, haematological and clinical studies of a silent beta-gene C → G mutation at 6 bp 3' to the termination codon (+1480 C → G) in twelve Greek families. *Br J Haematol* 103:45–51
- Mata J, Marguerat S, Bahler J (2005) Post-transcriptional control of gene expression: a genome-wide perspective. *Trends Biochem Sci* 30:506–514
- Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 288:911–940
- Medina P, Navarro S, Estelles A, Vaya A, Woodhams B, Mira Y, Villa P, Migaud-Fressart M, Ferrando F, Aznar J, Bertina RM, Espana F (2004) Contribution of polymorphisms in the endothelial protein C receptor gene to soluble endothelial protein C receptor and circulating activated protein C levels, and thrombotic risk. *Thromb Haemost* 91:905–911
- Medina P, Navarro S, Estelles A, Vaya A, Bertina RM, Espana F (2005) Influence of the 4600A/G and 4678G/C polymorphisms in the endothelial protein C receptor (*EPCR*) gene on the risk of venous thromboembolism in carriers of factor V Leiden. *Thromb Haemost* 94:389–394
- Meehan HA, Connell GJ (2001) The hairpin loop but not the bulged C of the iron responsive element is essential for high affinity binding to iron regulatory protein-1. *J Biol Chem* 276:14791–14796
- Meinsma D, Holthuisen PE, Van den Brande JL, Sussenbach JS (1991) Specific endonucleolytic cleavage of *IGF-II* mRNAs. *Biochem Biophys Res Commun* 179:1509–1516
- Meinsma D, Scheper W, Holthuisen PE, Van den Brande JL, Sussenbach JS (1992) Site-specific cleavage of *IGF-II* mRNAs requires sequence elements from two distinct regions of the *IGF-II* gene. *Nucleic Acids Res* 20:5003–5009
- Merkel M, Heeren J (2005) Give me A5 for lipoprotein hydrolysis!. *J Clin Invest* 115:2694–2696
- Meyer IM, Miklos I (2005) Statistical evidence for conserved, local secondary structure in the coding regions of eukaryotic mRNAs and pre-mRNAs. *Nucleic Acids Res* 33:6338–6348
- Meyre D, Bouatia-Naji N, Tounian A, Samson C, Lecoeur C, Vatin V, Ghossaini M, Wachter C, Hercberg S, Charpentier G, Patsch W, Pattou F, Charles MA, Tounian P, Clement K, Jouret B, Weill J, Maddux BA, Goldfine ID, Walley A, Boutin P, Dina C, Froguel P (2005) Variants of *ENPP1* are associated with childhood and adult obesity and increase the risk of glucose intolerance and type 2 diabetes. *Nat Genet* 37:863–867
- Mickleburgh I, Burtle B, Nury D, Chabanon H, Chrzanowska-Lightowlers Z, Hesketh JE (2004) Isolation and identification of a protein binding to the localization element of metallothionein-1 mRNA. *Biochem Soc Trans* 32:705–706
- Mickleburgh I, Burtle B, Hollas H, Campbell G, Chrzanowska-Lightowlers Z, Vedeler A, Hesketh J (2005) Annexin A2 binds to the localization signal in the 3' untranslated region of *c-myc* mRNA. *FEBS J* 272:413–421
- Mignone F, Gissi C, Liuni S, Pesole G (2002) Untranslated regions of mRNAs. *Genome Biol* 3:reviews0004.1–0004.10
- Miwa Y, Takiuchi S, Kamide K, Yoshii M, Horio T, Tanaka C, Banno M, Miyata T, Sasaguri T, Kawano Y (2004) Identification of gene polymorphism in lipocalin-type prostaglandin D synthase and its association with carotid atherosclerosis in Japanese hypertensive patients. *Biochem Biophys Res Commun* 322:428–433
- Morahan G, Huang D, Ymer SI, Cancilla MR, Stephen K, Dabadghao P, Werther G, Tait BD, Harrison LC, Colman PG (2001) Linkage disequilibrium of a type 1 diabetes susceptibility locus with a regulatory *IL12B* allele. *Nat Genet* 27:218–221
- Moreira A, Wollerton M, Monks J, Proudfoot NJ (1995) Upstream sequence elements enhance poly(A) site efficiency of the C2 complement gene and are phylogenetically conserved. *EMBO J* 14:3809–3819
- Moreira A, Takagaki Y, Brackenridge S, Wollerton M, Manley JL, Proudfoot NJ (1998) The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. *Genes Dev* 12:2522–2534
- Mori Y, Imaizumi K, Katayama T, Yoneda T, Tohyama M (2000) Two *cis*-acting elements in the 3' untranslated region of alpha-CaMKII regulate its dendritic targeting. *Nat Neurosci* 3:1079–1084
- Morohashi K, Takada T, Omori K, Suzuki E, Gejyo F (2003) Vascular endothelial growth factor gene polymorphisms in Japanese patients with sarcoidosis. *Chest* 123:1520–1526
- Mousses S, Ozelik H, Lee PD, Malkin D, Bull SB, Andrusis IL (1995) Two variants of the *CIP1/WAF1* gene occur together and are associated with human cancer. *Hum Mol Genet* 4:1089–1092
- Muhl A, Moslinger D, Item CB, Stockler-Ipsiroglu S (2001) Molecular characterisation of 34 patients with biotinidase deficiency ascertained by newborn screening and family investigation. *Eur J Hum Genet* 9:237–243
- Mueller T, Mas-Marques A, Sarrazin C, Wiese M, Halangk J, Witt H, Ahlenstiel G, Spengler U, Goebel U, Wiedenmann B, Schreier E, Berg T (2004) Influence of interleukin 12B (*IL12B*) polymorphisms on spontaneous and treatment-induced recovery from hepatitis C virus infection. *J Hepatol* 41:652–658

- Mummidi S, Ahuja SS, Gonzalez E, Anderson SA, Santiago EN, Stephan KT, Craig FE, O'Connell P, Tryon V, Clark RA, Dolan MJ, Ahuja SK (1998) Genealogy of the *CCR5* locus and chemokine system gene variants associated with altered rates of HIV-1 disease progression. *Nat Med* 4:786–793
- Munro TP, Magee RJ, Kidd GJ, Carson JH, Barbarese E, Smith LM, Smith R (1999) Mutational analysis of a heterogeneous nuclear ribonucleoprotein A2 response element for RNA trafficking. *J Biol Chem* 274:34389–34395
- Nakata K, Ujike H, Sakai A, Takaki M, Imamura T, Tanaka Y, Kuroda S (2003) The human dihydropyrimidinase-related protein 2 gene on chromosome 8p21 is associated with paranoid-type schizophrenia. *Biol Psychiatry* 53:571–576
- Natalizio BJ, Muniz LC, Arhin GK, Wilusz J, Lutz CS (2002) Upstream elements present in the 3'-untranslated region of collagen genes influence the processing efficiency of overlapping polyadenylation signals. *J Biol Chem* 277:42733–42740
- Nicaud V, Poirier O, Behague I, Herrmann SM, Mallet C, Troesch A, Bouyer J, Evans A, Luc G, Ruidavets JB, Arveiler D, Bingham A, Tiret L, Cambien F (1999) Polymorphisms of the endothelin-A and -B receptor genes in relation to blood pressure and myocardial infarction: the Etude Cas-Témoins sur l'Infarctus du Myocarde (ECTIM) Study. *Am J Hypertens* 12:304–310
- Niino M, Kikuchi S, Fukazawa T, Yabe I, Tashiro K (2003) Genetic polymorphisms of osteopontin in association with multiple sclerosis in Japanese patients. *J Neuroimmunol* 136:125–129
- Nury D, Chabanon H, Levadoux-Martin M, Hesketh J (2005) An eleven nucleotide section of the 3'-untranslated region is required for perinuclear localization of rat metallothionein-1 mRNA. *Biochem J* 387:419–428
- Oakley RH, Sar M, Cidlowski JA (1996) The human glucocorticoid receptor beta isoform. Expression, biochemical properties, and putative function. *J Biol Chem* 271:9550–9559
- O'Dell SD, Miller GJ, Cooper JA, Hindmarsh PC, Pringle PJ, Ford H, Humphries SE, Day IN (1997) Apal polymorphism in insulin-like growth factor II (*IGF2*) gene and weight in middle-aged males. *Int J Obes Relat Metab Disord* 21:822–825
- Ono K, Mannami T, Baba S, Tomoike H, Suga S, Iwai N (2002) A single-nucleotide polymorphism in C-type natriuretic peptide gene may be associated with hypertension. *Hypertens Res* 25:727–730
- de Pagter-Holthuizen P, Jansen M, van der Kammen RA, van Schaik FM, Sussenbach JS (1988) Differential expression of the human insulin-like growth factor II gene. Characterization of the *IGF-II* mRNAs and an mRNA encoding a putative IGF-II-associated protein. *Biochim Biophys Acta* 950:282–295
- Panza F, D'Introno A, Colacicco AM, Capurso C, Basile AM, Torres F, Capurso A, Solfrizzi V (2004) LBP-1c/CP2/LSF gene polymorphism and risk of sporadic Alzheimer's disease. *J Neurol Neurosurg Psychiatr* 75:166–168
- Peeters RP, van Toor H, Klootwijk W, de Rijke YB, Kuiper GG, Uitterlinden AG, Visser TJ (2003) Polymorphisms in thyroid hormone pathway genes are associated with plasma TSH and iodothyronine levels in healthy subjects. *J Clin Endocrinol Metab* 88:2880–2888
- Peeters RP, van den Beld AW, van Toor H, Uitterlinden AG, Janssen JA, Lamberts SW, Visser TJ (2005) A polymorphism in type I deiodinase is associated with circulating free insulin-like growth factor I levels and body composition in humans. *J Clin Endocrinol Metab* 90:256–263
- Pennacchio LA, Olivier M, Hubacek JA, Cohen JC, Cox DR, Fruchart JC, Krauss RM, Rubin EM (2001) An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science* 294:169–173
- Pennacchio LA, Olivier M, Hubacek JA, Krauss RM, Rubin EM, Cohen JC (2002) Two independent apolipoprotein A5 haplotypes influence human plasma triglyceride levels. *Hum Mol Genet* 11:3031–3038
- Peyvandi F, Garagiola I, Palla R, Marziliano N, Mannucci PM (2005) Role of the 2 adenine (g.11293_11294insAA) insertion polymorphism in the 3' untranslated region of the factor VII (*FVII*) gene: molecular characterization of a patient with severe FVII deficiency. *Hum Mutat* 26:455–461
- Pieczek M, Wax S, Beck AR, Kedersha N, Gupta M, Maritim B, Chen S, Gueydan C, Krus V, Streuli M, Anderson P (2000) TIA-1 is a translational silencer that selectively regulates the expression of TNF-alpha. *EMBO J* 19:4154–4163
- Pillai RS (2005) MicroRNA function: multiple mechanisms for a tiny RNA? *RNA* 11:1753–1761
- Pitzalis MV, Sarzani R, Dessi-Fulgheri P, Iacoviello M, Forleo C, Lucarelli K, Pietrucci F, Salvi F, Sorrentino S, Romito R, Guida P, Rappelli A, Rizzon P (2003) Allelic variants of natriuretic peptide receptor genes are associated with family history of hypertension and cardiovascular phenotype. *J Hypertens* 21:1491–1496
- Preuss CV, Wood TC, Szumlanski CL, Raftogianis RB, Otterness DM, Girard B, Scott MC, Weinshilboum RM (1998) Human histamine N-methyltransferase pharmacogenetics: common genetic polymorphisms that alter activity. *Mol Pharmacol* 53:708–717
- Puga I, Lainez B, Fernandez-Real JM, Buxade M, Broch M, Vendrell J, Espel E (2005) A polymorphism in the 3' untranslated region of the gene for tumor necrosis factor receptor 2 modulates reporter gene expression. *Endocrinology* 146:2210–2220
- Putland RA, Sassinis TA, Harvey JS, Diamond P, Coles LS, Brown CY, Goodall GJ (2002) RNA destabilization by the granulocyte colony-stimulating factor stem-loop destabilizing element involves a single stem-loop that promotes deadenylation. *Mol Cell Biol* 22:1664–1673
- Qian X, Shu A, Qin W, Xing Q, Gao J, Yang J, Feng G, He L (2004) A novel insertion mutation in the *FOXL2* gene is detected in a big Chinese family with blepharophimosis-ptosis-epicanthus inversus. *Mutat Res* 554:19–22
- van Rij RP, Broersen S, Goudsmit J, Coutinho RA, Schuitmaker H (1998) The role of a stromal cell-derived factor-1 chemokine gene variant in the clinical course of HIV-1 infection. *AIDS* 12:F85–F90
- Reddy KK, Oitomen FM, Patel GP, Bag J (2005) Perinuclear localization of slow troponin C mRNA in muscle cells is controlled by a cis-element located at its 3' untranslated region. *RNA* 11:294–307
- Robins H, Li Y, Padgett RW (2005) Incorporating structure to predict microRNA targets. *Proc Natl Acad Sci USA* 102:4006–4009
- Ross J (1995) mRNA stability in mammalian cells. *Microbiol Rev* 59:423–450
- Ross AF, Oleynikov Y, Kislauskis EH, Taneja KL, Singer RH (1997) Characterization of a beta-actin mRNA zipcode-binding protein. *Mol Cell Biol* 17:2158–2165
- Roth SM, Schrage MA, Metter EJ, Riechman SE, Fleg JL, Hurley BF, Ferrell RE (2002) *IGF2* genotype and obesity in men and women across the adult age span. *Int J Obes Relat Metab Disord* 26:585–587

- Rousseau D, Kaspar R, Rosenwald I, Gehrke L, Sonenberg N (1996) Translation initiation of ornithine decarboxylase and nucleocytoplasmic transport of cyclin D1 mRNA are increased in cells overexpressing eukaryotic initiation factor 4E. *Proc Natl Acad Sci USA* 93:1065–1070
- Ruggiero T, Olivero M, Follenzi A, Naldini L, Calogero R, Di Renzo MF (2003) Deletion in a (T)8 microsatellite abrogates expression regulation by 3'-UTR. *Nucleic Acids Res* 31:6561–6569
- Russell AI, Cunningham Graham DS, Shepherd C, Robertson CA, Whittaker J, Meeks J, Powell RJ, Isenberg DA, Walport MJ, Vyse TJ (2004) Polymorphism at the C-reactive protein locus influences gene expression and predisposes to systemic lupus erythematosus. *Hum Mol Genet* 13:137–147
- Sachchithanathan M, Stasinopoulos SJ, Wilusz J, Medcalf RL (2005) The relationship between the prothrombin upstream sequence element and the G20210A polymorphism: the influence of a competitive environment for mRNA 3'-end formation. *Nucleic Acids Res* 33:1010–1020
- Sakano S, Berggren P, Kumar R, Steineck G, Adolfsson J, Onelov E, Hemminki K, Larsson P (2003) Clinical course of bladder neoplasms and single nucleotide polymorphisms in the *CDKN2A* gene. *Int J Cancer* 104:98–103
- Santana da Silva LC, Carvalho TS, da Silva FB, Morari L, Fachel AA, Pires R, Refosco LF, Desnick RJ, Giugliani R, Saraiva Pereira ML (2003) Molecular characterization of phenylketonuria in South Brazil. *Mol Genet Metab* 79:17–24
- Schaaf MJ, Cidlowski JA (2002) AUUUA motifs in the 3'UTR of human glucocorticoid receptor alpha and beta mRNA destabilize mRNA and decrease receptor protein expression. *Steroids* 67:627–636
- Scheper W, Meinsma D, Holthuisen PE, Sussenbach JS (1995) Long-range RNA interaction of two sequence elements required for endonucleolytic cleavage of human insulin-like growth factor II mRNAs. *Mol Cell Biol* 15:235–245
- Schrager MA, Roth SM, Ferrell RE, Metter EJ, Russek-Cohen E, Lynch NA, Lindle RS, Hurley BF (2004) Insulin-like growth factor-2 genotype, fat-free mass, and muscle performance across the adult life span. *J Appl Physiol* 97:2176–2183
- Seegers D, Zwiers A, Strober W, Pena AS, Bouma G (2002) A *TaqI* polymorphism in the 3'UTR of the IL-12 p40 gene correlates with increased IL-12 secretion. *Genes Immun* 3:419–423
- Sgourou A, Papachatzopoulou A, Psiouri L, Antoniou M, Zoumbos N, Gibbs R, Athanassiadou A (2002) The beta-globin C → G mutation at 6 bp 3' to the termination codon causes beta-thalassaemia by decreasing the mRNA level. *Br J Haematol* 118:671–676
- Shao Y, Ismail-Beigi F (2004) Control of Na⁺-K⁺-ATPase beta 1-subunit expression: role of 3'-untranslated region. *Am J Physiol Cell Physiol* 286:C580–C585
- Sharma P, Hingorani A, Jia H, Hopper R, Brown MJ (1999) Quantitative association between a newly identified molecular variant in the endothelin-2 gene and human essential hypertension. *J Hypertens* 17:1281–1287
- Shav-Tal Y, Singer RH (2005) RNA localization. *J Cell Sci* 118:4077–4081
- Shaw G, Kamen R (1986) A conserved AU sequence from the 3' untranslated region of *GM-CSF* mRNA mediates selective mRNA degradation. *Cell* 46:659–667
- Shi J, Tian J, Pritchard A, Lendon C, Lambert JC, Iwatsubo T, Mann DM (2006) A 3'-UTR polymorphism in the oxidized LDL receptor 1 gene increases Abeta(40) load as cerebral amyloid angiopathy in Alzheimer's disease. *Acta Neuropathol* 111:15–20
- Shiffman D, Ellis SG, Rowland CM, Malloy MJ, Luke MM, Iakoubova OA, Pullinger CR, Cassano J, Aouizerat BE, Fenwick RG, Reitz RE, Catanese JJ, Leong DU, Zellner C, Sninsky JJ, Topol EJ, Devlin JJ, Kane JP (2005) Identification of four gene variants associated with myocardial infarction. *Am J Hum Genet* 77:596–605
- Small KM, Miale-Perez J, Seman CA, Theiss CT, Brown KM, Liggett SB (2004) Polymorphisms of cardiac presynaptic alpha2C adrenergic receptors: Diverse intragenic variability with haplotype-specific functional effects. *Proc Natl Acad Sci USA* 101:13020–13025
- Spencer-Jones NJ, Wang X, Snieder H, Spector TD, Carter ND, O'dell SD (2005) Protein tyrosine phosphatase-1B gene *PTPNI*: selection of tagging single nucleotide polymorphisms and association with body fat, insulin sensitivity, and the metabolic syndrome in a normal female population. *Diabetes* 54:3296–3304
- Spurdle AB, Hopper JL, Chen X, McCredie MR, Giles GG, Newman B, Chenevix-Trench G (2002) Prohibitin 3' untranslated region polymorphism and breast cancer risk in Australian women. *Lancet* 360:925–926
- Spurdle AB, Purdie DM, Chen X, Chenevix-Trench G (2003) The prohibitin 3' untranslated region polymorphism is not associated with risk of ovarian cancer. *Gynecol Oncol* 90:145–149
- Stanilova S, Miteva L (2005) *TaqI* polymorphism in 3'UTR of the *IL-12B* and association with IL-12p40 production from human PBMC. *Genes Immun* 6:364–366
- Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, Abeyasinghe S, Krawczak M, Cooper DN (2003) Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat* 21:577–581
- Stoeckle MY (1992) Removal of a 3' non-coding sequence is an initial step in degradation of gro alpha mRNA and is regulated by interleukin-1. *Nucleic Acids Res* 20:1123–1127
- Stoecklin G, Lu M, Rattenbacher B, Moroni C (2003) A constitutive decay element promotes tumor necrosis factor alpha mRNA degradation via an AU-rich element-independent pathway. *Mol Cell Biol* 23:3506–3515
- Styrkarsdottir U, Cazier JB, Kong A, Rolfsson O, Larsen H, Bjarnadottir E, Johannsdottir VD, Sigurdardottir MS, Bagger Y, Christiansen C, Reynisdottir I, Grant SF, Jonasson K, Frigge ML, Gulcher JR, Sigurdsson G, Stefansson K (2003) Linkage of osteoporosis to chromosome 20p12 and association to *BMP2*. *PLoS Biol* 1:E69
- Suviolahti E, Ridderstrale M, Almgren P, Klannemark M, Melander O, Carlsson E, Carlsson M, Hedenbro J, Orholm Melander M (2003) Pro-opiomelanocortin gene is associated with serum leptin levels in lean but not in obese individuals. *Int J Obes Relat Metab Disord* 27:1204–1211
- Tadokoro K, Fujii H, Inoue T, Yamada M (1991) Polymerase chain reaction (PCR) for detection of *ApaI* polymorphism at the insulin like growth factor II gene (*IGF2*). *Nucleic Acids Res* 19:6967
- Talmud PJ, Palmen J, Putt W, Lins L, Humphries SE (2005) Determination of the functionality of common *APOA5* polymorphisms. *J Biol Chem* 280:28215–28220
- Tamminen M, Kakko S, Kesaniemi YA, Savolainen MJ (1996) A polymorphic site in the 3' untranslated region of the cholesterol ester transfer protein (*CETP*) gene is associated with low CETP activity. *Atherosclerosis* 124:237–247
- Tan MS, Chang SY, Chang DM, Tsai JC, Lee YJ (2003) Association of resistin gene 3'-untranslated region +62G → A polymorphism with type 2 diabetes and hypertension in a Chinese population. *J Clin Endocrinol Metab* 88:1258–1263

- Taylor AE, Yip A, Brayne C, Easton D, Evans JG, Xuereb J, Cairns N, Esiri MM, Rubinsztein DC (2001) Genetic association of an LBP-1c/CP2/LSF gene polymorphism with late onset Alzheimer's disease. *J Med Genet* 38:232–233
- Trabetti E, Biscuola M, Cavallari U, Malerba G, Girelli D, Olivieri O, Martinelli N, Corrocher R, Pignatti PF (2006) On the association of the oxidised LDL receptor 1 (*OLR1*) gene in patients with acute myocardial infarction or coronary artery disease. *Eur J Hum Genet* 14:127–130
- Tran H, Schilling M, Wirbelauer C, Hess D, Nagamine Y (2004) Facilitation of mRNA deadenylation and decay by the exosome-bound, DEXH protein RHAU. *Mol Cell* 13:101–111
- Tsai FJ, Wu HC, Chen HY, Lu HF, Hsu CD, Chen WC (2003) Association of E-cadherin gene 3'-UTR C/T polymorphism with calcium oxalate stone disease. *Urol Int* 70:278–281
- Ulrich CM, Bigler J, Velicer CM, Greene EA, Farin FM, Potter JD (2000) Searching expressed sequence tag databases: discovery and confirmation of a common polymorphism in the thymidylate synthase gene. *Cancer Epidemiol Biomarkers Prev* 9:1381–1385
- Valdes AM, Hart DJ, Jones KA, Surdulescu G, Swarbrick P, Doyle DV, Schafer AJ, Spector TD (2004) Association study of candidate genes for the prevalence and progression of knee osteoarthritis. *Arthritis Rheum* 50:2497–2507
- Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R (2006) Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev* 20:515–524
- Vasilopoulos Y, Cork MJ, Murphy R, Williams HC, Robinson DA, Duff GW, Ward SJ, Tazi-Ahmini R (2004) Genetic association between an AACC insertion in the 3'UTR of the stratum corneum chymotryptic enzyme gene and atopic dermatitis. *J Invest Dermatol* 123:62–66
- Vielhaber E, Jacobson DP, Ketterling RP, Liu JZ, Sommer SS (1993) A mutation in the 3' untranslated region of the factor IX gene in four families with hemophilia B. *Hum Mol Genet* 2:1309–1310
- Villette S, Kyle JA, Brown KM, Pickard K, Milne JS, Nicol F, Arthur JR, Hesketh JE (2002) A novel single nucleotide polymorphism in the 3' untranslated region of human glutathione peroxidase 4 influences lipoxigenase metabolism. *Blood Cells Mol Dis* 29:174–178
- Waggoner SA, Liebhaber SA (2003) Regulation of alpha-globin mRNA stability. *Exp Biol Med* 228:387–395
- Walczak R, Westhof E, Carbon P, Krol A (1996) A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *RNA* 2:367–379
- Walczak R, Carbon P, Krol A (1998) An essential non-Watson-Crick base pair motif in 3'UTR to mediate selenoprotein translation. *RNA* 4:74–84
- Wang Z, Kiledjian M (2000) Identification of an erythroid-enriched endoribonuclease activity involved in specific mRNA cleavage. *EMBO J* 19:295–305
- Wang WY, Zee RY, Morris BJ (1997) Association of angiotensin II type 1 receptor gene polymorphism with essential hypertension. *Clin Genet* 51:31–34
- Wang Z, Day N, Trifillis P, Kiledjian M (1999) An mRNA stability complex functions with poly(A)-binding protein to stabilize mRNA *in vitro*. *Mol Cell Biol* 19:4552–4560
- Wang J, Pitarque M, Ingelman-Sundberg M (2006) 3'-UTR polymorphism in the human *CYP2A6* gene affects mRNA stability and enzyme expression. *Biochem Biophys Res Commun* 340:491–497
- Webb KE, Martin JF, Hamsten A, Eriksson P, Iacoviello L, Gattone M, Donati MB, Di Castelnuovo A, Erusalimsky J, Humphries SE (2001) Polymorphisms in the thrombopoietin gene are associated with risk of myocardial infarction at a young age. *Atherosclerosis* 154:703–711
- Weiss IM, Liebhaber SA (1994) Erythroid cell-specific determinants of alpha-globin mRNA stability. *Mol Cell Biol* 14:8123–8132
- Weiss IM, Liebhaber SA (1995) Erythroid cell-specific mRNA stability elements in the alpha 2-globin 3' nontranslated region. *Mol Cell Biol* 15:2457–2465
- Windsor L, Morahan G, Huang D, McCann V, Jones T, James I, Christiansen FT, Price P (2004) Alleles of the *IL12B* 3'UTR associate with late onset of type 1 diabetes. *Hum Immunol* 65:1432–1436
- Winkler C, Modi W, Smith MW, Nelson GW, Wu X, Carrington M, Dean M, Honjo T, Tashiro K, Yabe D, Buchbinder S, Vittinghoff E, Goedert JJ, O'Brien TR, Jacobson LP, Detels R, Donfield S, Willoughby A, Gomperts E, Vlahov D, Phair J, ALIVE Study, Hemophilia Growth, Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC), O'Brien SJ (1998) Genetic restriction of AIDS pathogenesis by an *SDF-1* chemokine gene variant. *Science* 279:389–393
- Wolford JK, Hanson RL, Kobes S, Bogardus C, Prochazka M (2001) Analysis of linkage disequilibrium between polymorphisms in the *KCNJ9* gene with type 2 diabetes mellitus in Pima Indians. *Mol Genet Metab* 73:97–103
- Wolford JK, Colligan PB, Gruber JD, Bogardus C (2003) Variants in the interleukin 6 receptor gene are associated with obesity in Pima Indians. *Mol Genet Metab* 80:338–343
- Wu HC, Lai MT, Wu CI, Chen HY, Wan L, Tsai FJ, Chen WC (2005) E-cadherin gene 3'-UTR C/T polymorphism is associated with prostate cancer. *Urol Int* 75:350–353
- Xia J, Scherer SW, Cohen PT, Majer M, Xi T, Norman RA, Knowler WC, Bogardus C, Prochazka M (1998) A common variant in *PPPIR3* associated with insulin resistance and type 2 diabetes. *Diabetes* 47:1519–1524
- Xia J, Bogardus C, Prochazka M (1999) A type 2 diabetes-associated polymorphic ARE motif affecting expression of *PPPIR3* is involved in RNA-protein interactions. *Mol Genet Metab* 68:48–55
- Xiang K, Cox NJ, Bell GI (1988) *ApaI* and *SstI* RFLPs at the insulin-like growth factor II (*IGF2*) locus on chromosome 11. *Nucleic Acids Res* 16:3599
- Yang MS, Yu L, Guo TW, Zhu SM, Liu HJ, Shi YY, Gu NF, Feng GY, He L (2004) Evidence for association between single nucleotide polymorphisms in T complex protein 1 gene and schizophrenia in the Chinese Han population. *J Med Genet* 41:e63
- Yilmaz V, Yentur SP, Saruhan-Direskeneli G (2005) IL-12 and IL-10 polymorphisms and their effects on cytokine production. *Cytokine* 30:188–194
- Yin LM, Zhu WF, Wei L, Xu XY, Sun DG, Wang YB, Fan WM, Yu M, Tian XL, Wang QX, Gao Y, Zhuang H (2004) Association of interleukin-12 p40 gene 3'-untranslated region polymorphism and outcome of HCV infection. *World J Gastroenterol* 10:2330–2333
- Yu J, Russell JE (2001) Structural and functional analysis of an mRNA complex that mediates the high stability of human beta-globin mRNA. *Mol Cell Biol* 21:5879–5888
- Zehner ZE, Shepherd RK, Gabryszuk J, Fu TF, Al-Ali M, Holmes WM (1997) RNA-protein interactions within the 3' untranslated region of vimentin mRNA. *Nucleic Acids Res* 25:3362–3370

- Zhang H, Jia Y, Cooper JJ, Hale T, Zhang Z, Elbein SC (2004) Common variants in glutamine:fructose-6-phosphate amidotransferase 2 (*GFPT2*) gene are associated with type 2 diabetes, diabetic nephropathy, and increased *GFPT2* mRNA levels. *J Clin Endocrinol Metab* 89:748–755
- Zhang Z, Xu Y, Zhou J, Wang X, Wang L, Hu X, Guo J, Wei Q, Shen H (2005) Polymorphisms of thymidylate synthase in the 5'- and 3'-untranslated regions associated with risk of gastric cancer in South China: a case-control analysis. *Carcinogenesis* 26:1764–1769
- Zhao J, Hyman L, Moore C (1999) Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev* 63:405–445
- Zhao Y, Samal E, Srivastava D (2005) Serum response factor regulates a muscle-specific microRNA that targets *Hand2* during cardiogenesis. *Nature* 436:214–220
- Zhou XL, Eriksson U, Werelius B, Kressner U, Sun XF, Lindblom A (2004) Definition of candidate low risk *APC* alleles in a Swedish population. *Int J Cancer* 110:550–555
- Zuker M (2003) *Mfold* web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406–3415