ON THE COMPOSITION OF HYPER-CONSERVATION SCORE (HCS)

We defined the sequence conservation measure, which we call Hyper Conservation Score (HCS), by first selecting one of the two conservation measures defined for the 44-way alignments available at the UCSC genome browser (33). We choose as Sequence Conservation Score (SCS) the *phastCons*-derived metric (34) instead of the *phyloP* one (105), as the former considers neighboring bases in determining a base score, being thus sensible to stretches of conserved bases: this fact makes it more suitable for identifying conserved elements than *phyloP*, which instead computes conservation independently at each position. *phastCons* takes into account the phylogenetic tree to estimate the probabilities for bases to be conserved or not in the HMM models it is based upon. Nevertheless, being our aim to identify exceptionally conserved sequence stretches because of their potential functional meaning as cis components of core posttranscriptional networks, we estimated as essential the requirement for sharing of the sequences among the different vertebrate species considered. To put more weight on the phylogenetic distance, we included in our metric the Branch Length Score (BLS) as introduced in a comparison between close *Drosophila* species (35). This measure is the proportion of the distance covered by the branches of the phylogenetic tree by the alignment of a particular sequence, thus giving more importance to elements conserved across a wide range of species than to the ones restricted to a group of closely related species. We argued that, while phylogenetic information are already included in SCS, BLS would have been not redundant. To verify this we computed the Pearson correlation coefficient between SCS and BLS, obtaining a value of 0.48, which indicates only a moderate correlation of the two components of our HCS. This result confirms that the BLS usefully complements the SCS.

We further had to find a convenient measure of relative weight of SCS and BLS in HCS. We performed several runs of our pipeline, varying the SCS-BLS score composition from SCS only (100%-0%) to BLS only (0%-100%), through five intermediate proportions (80%-20%; 60%-40%; 50%-50%; 40%-60%; 20%-80%). What we obtain as result is a progressive increase in HCE sizes in parallel with a marked reduction of their total number. While more than 120000 HCEs are produced in the first two runs (100%-

0%, 80%-20%), only 3149 are retained in the half-half proportion (50%-50%), and this number goes down to just 232 HCEs for the BLS-only run. Median and average HCE lengths increase respectively from 62 and 17 bases to 114 and 249 bases: the 50%-50% case has a median length of 23 bases and an average length of 100 bases. We selected the 50% SCS and 50% BLS composition as our final conservation measure, because of the number of selected HCEs identified a small percentage of the total UTR space (0.47%) and a corresponding small percentage of mRNAs (1.8%). With this choice we believed to have greatly reduced the number of false positives HCEs in our final dataset.

SUPPLEMENTARY REFERENCES

33. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M et al. (2011) The UCSC Genome Browser database: update 2011.Nucleic Acids Res., 39(suppl 1), D876-D882.

34. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res., 15(8), 1034-50.

35. Stark A, Lin MF, Kheradpour P, Pedersen JS, Parts L, Carlson JW, Crosby MA et al. (2007) Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures.Nature, 450(7167),219-232.

105. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res., 20(1), 110-21.

Gene symbol	Gene name	Uniprot description	RRM-only architecture
CPEB1	Cytoplasmic polyadenylation element-binding protein 1	Sequence-specific RNA-binding protein that regulates mRNA cytoplasmic polyadenylation and translation initiation during oocyte maturation, early development and at postsynapse sites of neurons. Binds to the cytoplasmic polyadenylation element (CPE), an uridine-rich sequence element (consensus sequence 5'-UUUUUAU-3') within the mRNA 3'-UTR. In absence of phosphorylation and in association with TACC3 is also involved as a repressor of translation of CPE-containing mRNA; a repression that is relieved by phosphorylation or degradation	**v**
CUGBP1	CUGBP Elav-like family member 1	RNA-binding protein implicated in the regulation of several post-transcriptional events. Involved in pre-mRNA alternative splicing, mRNA translation and stability. Mediates exon inclusion and/or exclusion in pre-mRNA that are subject to tissue-specific and developmentally regulated alternative splicing.	**v**
EIF4B	Eukaryotic translation initiation factor 4B	Required for the binding of mRNA to ribosomes. Functions in close association with EIF4-F	**v**

		and EIF4-A. Binds near the 5'-terminal cap of mRNA in presence of EIF-4F and ATP. Promotes the ATPase activity and the ATP-dependent RNA unwinding activity of both EIF4-A and EIF4-F.	
ELAVL4	ELAV-like protein 4	May play a role in neuron-specific RNA processing. Protects CDKN1A mRNA from decay by binding to its 3'-UTR. Binds to AU-rich sequences (AREs) of target mRNAs, including VEGF and FOS mRNA.	**v**
EWSR1	RNA-binding protein EWS	Might normally function as a repressor. EWS-fusion-proteins (EFPS) may play a role in the tumorigenic process. They may disturb gene expression by mimicking, or interfering with the normal function of CTD-POLII within the transcription initiation complex. They may also contribute to an aberrant activation of the fusion protein target genes.	**x**
FUS	RNA-binding protein FUS	Binds both single-stranded and double-stranded DNA and promotes ATP-independent annealing of complementary single-stranded DNAs and D-loop formation in superhelical double-stranded DNA. May play a role in maintenance of genomic integrity.	**x**
HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	Involved in the packaging of pre-mRNA into hnRNP particles, transport of	**v**

		poly(A) mRNA from the nucleus to the cytoplasm and may modulate splice site selection. May play a role in HCV RNA replication.	
HNRNPA3	Heterogeneous nuclear ribonucleoprotein A3	Plays a role in cytoplasmic trafficking of RNA. Binds to the cis-acting response element, A2RE. May be involved in pre-mRNA splicing.	**v**
HNRNPD	Heterogeneous nuclear ribonucleoprotein D0	Binds with high affinity to RNA molecules that contain AU-rich elements (AREs) found within the 3'-UTR of many proto-oncogenes and cytokine mRNAs. Also binds to double- and single-stranded DNA sequences in a specific manner and functions a transcription factor. Each of the RNA-binding domains specifically can bind solely to a single-stranded non-monotonous 5'-UUAG-3' sequence and also weaker to the single-stranded 5'-TTAGGG-3' telomeric DNA repeat. Binds RNA oligonucleotides with 5'-UUAGGG-3' repeats more tightly than the telomeric single-stranded DNA 5'-TTAGGG-3' repeats. Binding of RRM1 to DNA inhibits the formation of DNA quadruplex structure which may play a role in telomere elongation. May be involved in translationally coupled	**v**

		mRNA turnover. Implicated with other RNA-binding proteins in the cytoplasmic deadenylation/translational and decay interplay of the FOS mRNA mediated by the major coding-region determinant of instability (mCRD) domain.	
HNRNPM	Heterogeneous nuclear ribonucleoprotein M	Pre-mRNA binding protein in vivo, binds avidly to poly(G) and poly(U) RNA homopolymers in vitro. Involved in splicing. Acts as a receptor for carcinoembryonic antigen in Kupffer cells, may initiate a series of signaling events leading to tyrosine phosphorylation of proteins and induction of IL-1 alpha, IL-6, IL-10 and tumor necrosis factor alpha cytokines.	**v**
MSI2	RNA-binding protein Musashi homolog 2	RNA binding protein that regulates the expression of target mRNAs at the translation level. May play a role in the proliferation and maintenance of stem cells in the central nervous system	**v**
PABPC1	Polyadenylate-binding protein 1	Binds the poly(A) tail of mRNA. May be involved in cytoplasmic regulatory processes of mRNA metabolism such as pre-mRNA splicing. Its function in translational initiation regulation can either be enhanced by PAIP1 or repressed by PAIP2. Can probably bind	**x**

		to cytoplasmic RNA sequences other than poly(A) in vivo. May be involved in translationally coupled mRNA turnover. Implicated with other RNA-binding proteins in the cytoplasmic deadenylation/translational and decay interplay of the FOS mRNA mediated by the major coding-region determinant of instability (mCRD) domain.	
PPARGC1B	Peroxisome proliferator-activated receptor gamma coactivator 1-beta	Plays a role of stimulator of transcription factors and nuclear receptors activities. Activates transcritional activity of estrogen receptor alpha, nuclear respiratory factor 1 (NRF1) and glucocorticoid receptor in the presence of glucocorticoids. May play a role in constitutive non-adrenergic-mediated mitochondrial biogenesis as suggested by increased basal oxygen consumption and mitochondrial number when overexpressed. May be involved in fat oxidation and non-oxidative glucose metabolism and in the regulation of energy expenditure.	**v**
PTBLP	Polypyrimidine tract-binding protein 2	RNA-binding protein which binds to intronic polypyrimidine tracts and mediates negative regulation of exons splicing. May antagonize in a tissue-specific manner the ability of NOVA1 to	**v**

		activate exon selection. Beside its function in pre-mRNA splicing, plays also a role in the regulation of translation. Isoform 5 has a reduced affinity for RNA.	
RBM15	Putative RNA-binding protein 15	May be implicated in HOX gene regulation.	**x**
RBM16	Putative RNA-binding protein 16	May play a role in mRNA processing.	**x**
RBM26	RNA-binding protein 26		**x**
RBMS1	RNA-binding motif, single-stranded-interacting protein 1	Single-stranded DNA binding protein that interacts with the region upstream of the MYC gene. Binds specifically to the DNA sequence motif 5'-[AT]CT[AT][AT]T-3'. Probably has a role in DNA replication.	**v**
RBMS3	RNA-binding motif, single-stranded-interacting protein 3	Binds poly(A) and poly(U) oligoribonucleotides.	**v**
SFRS11	Splicing factor, arginine/serine-rich 11	May function in pre-mRNA splicing.	**v**
SFRS15	Splicing factor, arginine/serine-rich 15	May act to physically and functionally link transcription and pre-mRNA processing	**x**
SYNCRIP	synaptotagmin binding, cytoplasmic RNA interacting protein, hnRNPQ	Heterogenous nuclear ribonucleoprotein (hnRNP) implicated in mRNA processing mechanisms.Component of the CRD-mediated complex that promotes MYC mRNA stability.	**v**
THOC4	THO complex subunit 4	CRD-mediated complex that promotes MYC mRNA stability.	v

Supplementary Table 1: List of the 23 HCE-containing RRM-type RBP identified by our pipeline. Listed are gene symbol, name, Uniprot gene function description and whether the protein contains only RRM or also other domains.

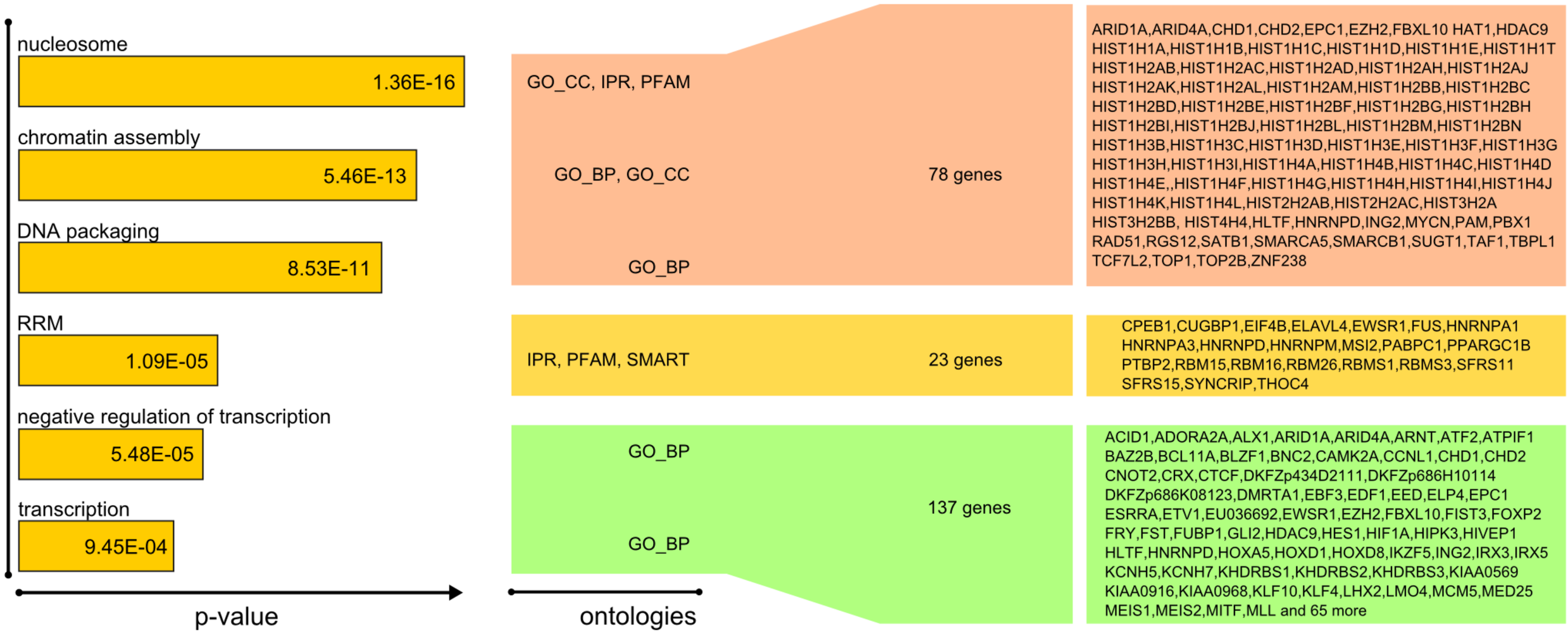

Figure S1. HCEs cluster in genes belonging to three different biological functions.

Ontology enrichment analysis of HCEs-containing genes highlights three groups of genes corresponding to three different biological functions. Multiple ontologies were used to infer possible functional groupings: the top results are a most significant group composed of genes involved in chromosomes assembly, a significant set consisting of 23 genes coding for RRM-containing genes for RBPs and a third, less significant group of genes playing a role in transcription. Here the ontology terms clusters giving rise to these groups are shown, along with their enrichment p-value and the final list of involved genes.

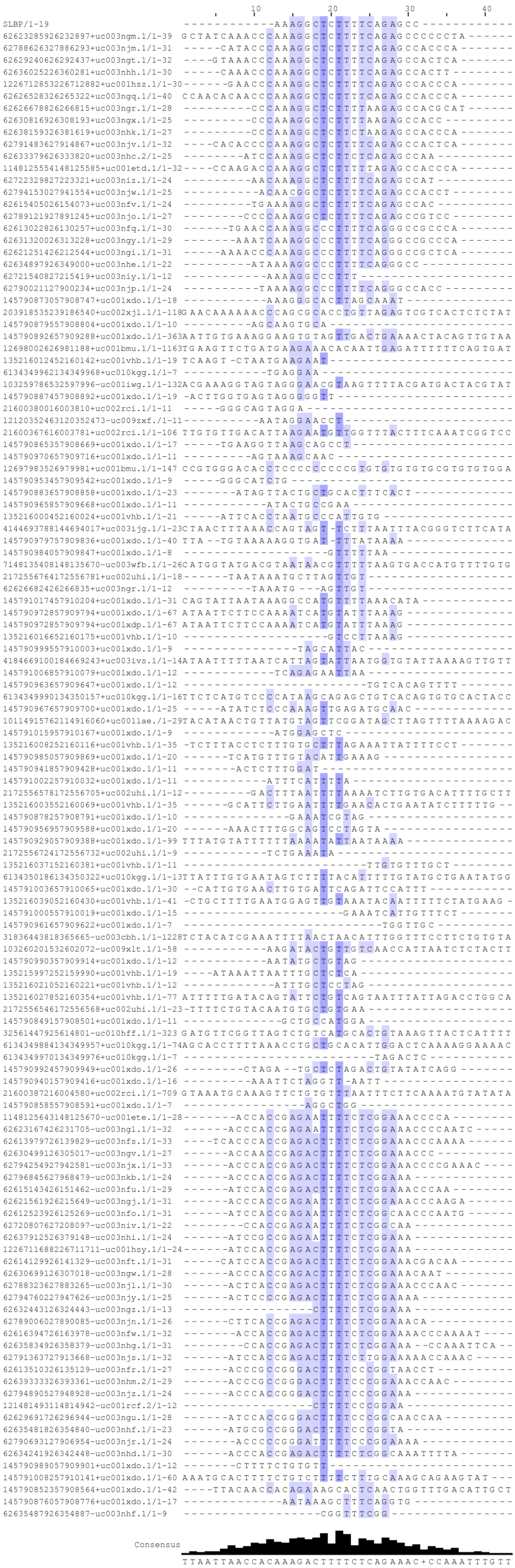

Figure S2. HCEs in 3'UTR of chromosome assembly genes identify SLBP binding sites.

A significant fraction of HCEs found in the 3'UTR of genes belonging to the chromosome assembly functional group was noticed to harbor a sequence corresponding to the binding motif of the stem-loop binding protein (SLBP), which is known to bind to the 3'UTR of histone genes and to stabilize the mRNA in order to compensate for the absence of a poly(A) tail. This stabilization mechanism is known to be heavily conserved and can thus be considered as a benchmark for our HCE identification method. Here the ClustalW2 alignment of these HCEs with the SLBP binding motif (the first sequence in the alignment) is displayed.

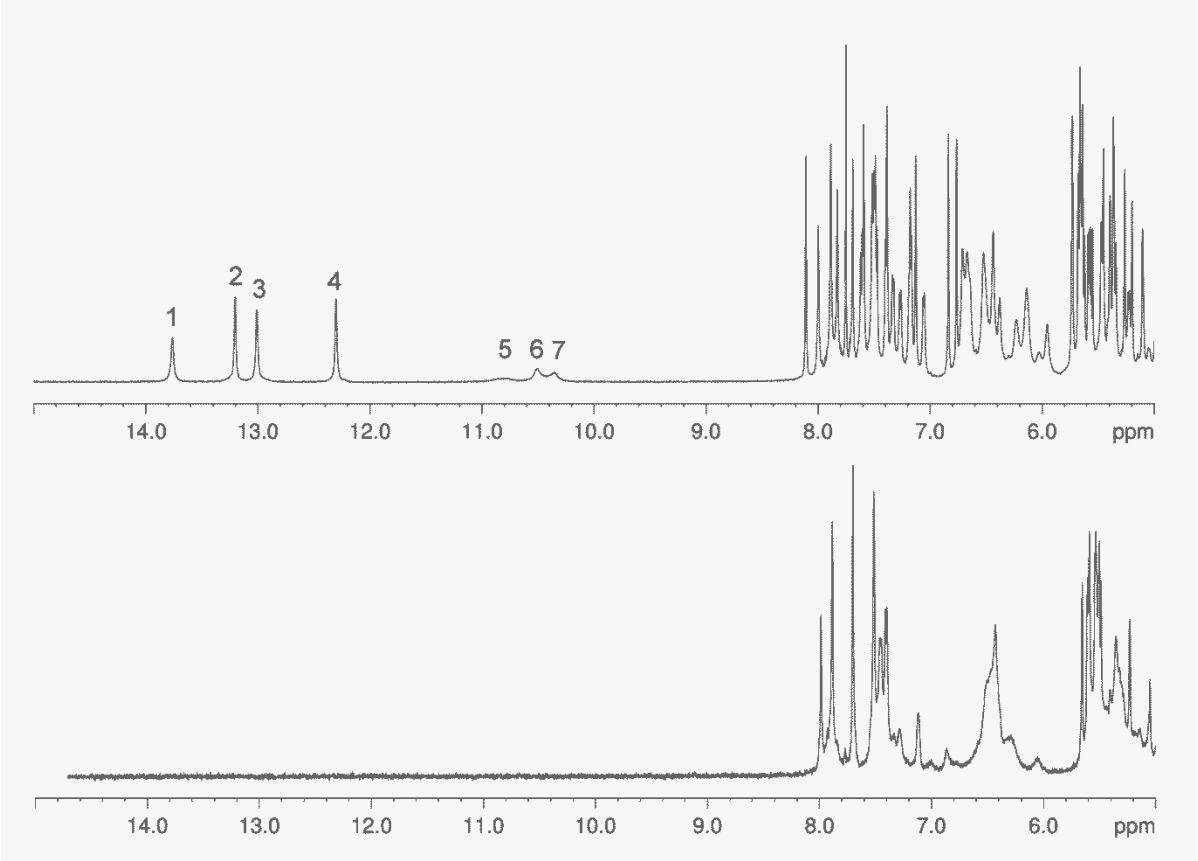

Figure S3. ^{1}H-NMR spectrum of SL (top) and NF (bottom). Only the downfield portion of the spectra are displayed for better visualization of the imino/amino-proton region. Spectra were acquired at 5 °C on 0.1 mM samples in 10 mM phosphate buffer, pH 7, and 10 mM NaCl. Imino proton peaks are labeled by numbers.

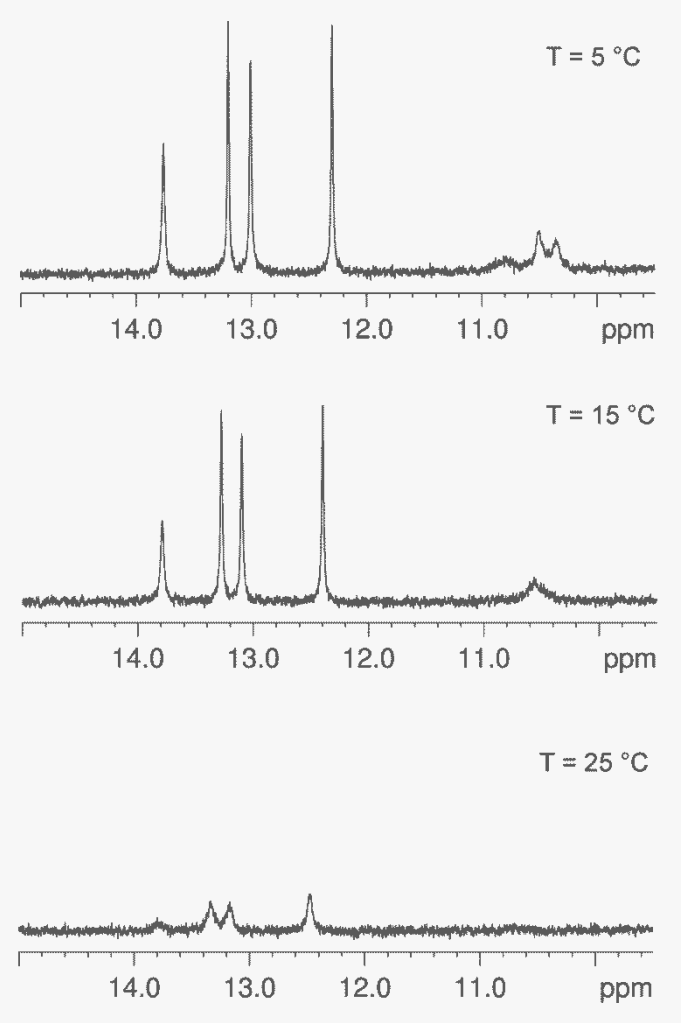

Figure S4. Temperature dependence of the ^{1}H-NMR spectrum of SL. Only the imino-proton region is displayed. Spectra were acquired with the same acquisition parameters on 0.1 mM samples in 10 mM phosphate buffer, pH 7, and 10 mM NaCl.

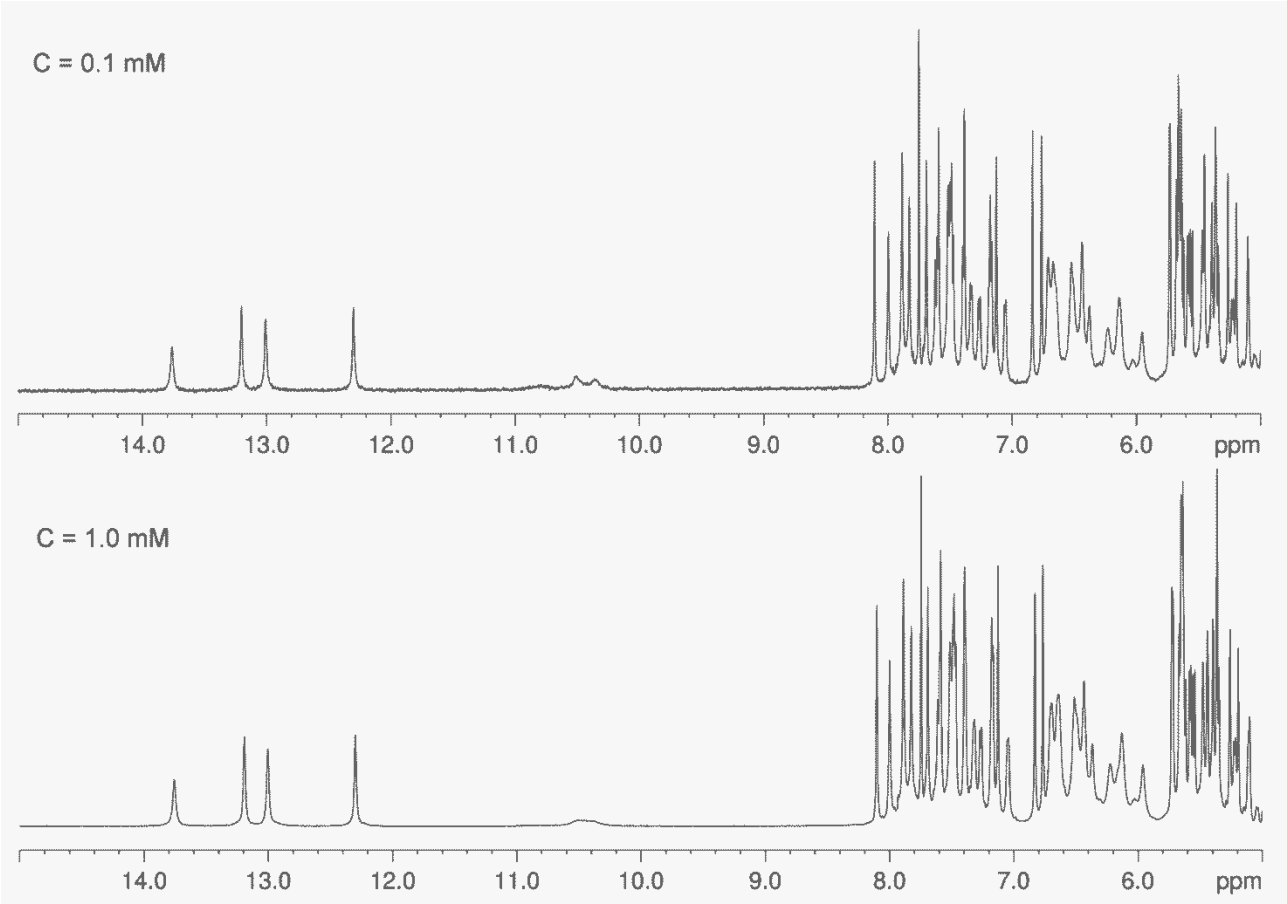

Figure S5. Concentration dependence of the ^{1}H-NMR spectrum of SL. Only the downfield portion of the spectra are displayed for better visualization of the imino/amino-proton region. Spectra were acquired at 5 °C on 0.1 mM (top) and 1.0 mM (bottom) samples in 10 mM phosphate buffer, pH 7, and 10 mM NaCl.