

# Evidence for a Quadruplex Structure in the Polymorphic hs1.2 Enhancer of the Immunoglobulin Heavy Chain 3' Regulatory Regions and its Conservation in Mammals

Marco Sette,<sup>1\*</sup> Pietro D'Addabbo,<sup>2\*</sup> Geoffrey Kelly,<sup>3</sup> Alessandro Cicconi,<sup>4,5</sup> Emanuela Micheli,<sup>4,5</sup> Stefano Cacchione,<sup>4,5</sup> Anna Poma,<sup>6</sup> Cesare Gargioli,<sup>7</sup> Vincenzo Giambra,<sup>8</sup> Domenico Frezza<sup>7</sup>

<sup>1</sup> Department of Chemical Sciences and Technology, University of Roma "Tor Vergata", Roma, Italy

<sup>2</sup> Department of Biology, University of Bari "Aldo Moro", Bari, Italy

<sup>3</sup> MRC Biomedical NMR Centre, The Francis Crick Institute, Mill Hill Laboratory, London, UK

<sup>4</sup> Department of Biology and Biotechnology, Sapienza University, Roma, Italy

<sup>5</sup> Institute Pasteur-Fondazione Cenci-Bolognetti, Roma, Italy

<sup>6</sup> Department of Life, Health and Environmental Sciences, University of L'Aquila, L'Aquila, Italy

<sup>7</sup> Department of Biology, University of Roma "Tor Vergata", Roma, Italy

<sup>8</sup> "Terry Fox" Laboratory, BC Cancer Agency, Vancouver, Canada

Received 10 April 2016; accepted 1 June 2016

Published online 10 June 2016 in Wiley Online Library ([wileyonlinelibrary.com](http://wileyonlinelibrary.com)). DOI 10.1002/bip.22891

## ABSTRACT:

Regulatory regions in the genome can act through a variety of mechanisms that range from the occurrence of histone modifications to the presence of protein-binding loci for self-annealing sequences. The final result is often the induction of a conformational change of the DNA double helix, which alters the accessibility of a region to transcription factors and consequently gene expression. A ~300 kb regulatory region on chromosome 14 at the 3' end (3'RR) of immunoglobulin (Ig) heavy-chain genes shows very peculiar features, conserved in mammals, including enhancers and transcription factor binding sites. In primates, the 3'RR is present in two copies, both having a central enhancer named hs1.2. We previously demonstrated the association between differ-

ent hs1.2 alleles and Ig plasma levels in immunopathology. Here, we present the analysis of a putative G-quadruplex structure (tetraplex) consensus site embedded in a variable number tandem repeat (one to four copies) of hs1.2 that is a distinctive element among the enhancer alleles, and an investigation of its three-dimensional structure using bioinformatics and spectroscopic approaches. We suggest that both the role of the enhancer and the alternative effect of the hs1.2 alleles may be achieved through their peculiar three-dimensional-conformational rearrangement. © 2017 The Authors Biopolymers Published by Wiley Periodicals, Inc. *Biopolymers* 105: 768–778, 2016.

**Keywords:** polymorphic enhancer hs1.2; circular dichroism; nuclear magnetic resonance; quadruplex DNA; immunoglobulins

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

The copyright line for this article was changed on 10th May 2017 after original online publication.

Correspondence to: Marco Sette; e-mail: [sette@uniroma2.it](mailto:sette@uniroma2.it) and Pietro D'Addabbo; e-mail: [pietro.daddabbo@uniba.it](mailto:pietro.daddabbo@uniba.it)

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© 2017 The Authors Biopolymers Published by Wiley Periodicals, Inc.

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version. You can request a copy of any preprints from the past two calendar years by emailing the Biopolymers editorial office at [biopolymers@wiley.com](mailto:biopolymers@wiley.com).

## INTRODUCTION

**A** relevant role of noncoding and intergenic regions in the human genome control has been recognized for many years. Emphasis has been given to the function of these regions in genetic regulation.<sup>1–3</sup> This task is achieved either through binding of specific proteins or through inducing specific DNA conformations, which result in the control of entire coding regions.<sup>4,5</sup>

We are interested in understanding how the 3' regulative region (3'RR) of immunoglobulin heavy chain gene cluster (IgH) works. During B-cell maturation, the 3'RR plays fundamental control roles for somatic hypermutation, and in class switch through epigenetic changes of chromatin.<sup>6,7</sup> The 3'RR is present in a single copy in nonprimate mammals, while is duplicated in human on chromosome 14, and in most primate species on related syntenic regions.<sup>8</sup> Despite synteny among 3'RRs in different mammals, these regions can harbor three enhancers (hs1.2, hs3, and hs4) as in humans, or four enhancers, one of the original three being duplicated, as in rodents, in which also three insulators (hs5, hs6, and hs7) are present in a lineage-specific sequence.<sup>9</sup> In all species, hs1.2 maps to the central conserved core of the 3'RRs. In the genomes harboring the duplication, binding sites for specific transcription factors like NF- $\kappa$ B and Sp1, are conserved in both the paralogous copies of the hs1.2 enhancer.<sup>8</sup> These sites and their peptidic interactors are thought to play a crucial role both in the isotopic switch and in correct Ig transcription at different stages of B-cell development.<sup>6,7,10–13</sup> In agreement with these findings, we demonstrated that human polymorphisms of hs1.2 enhancer in 3'RR1 are associated with the humoral-response control of Ig-circulating levels and with autoimmune diseases.<sup>14–16</sup> Moreover, an imbalance of allele frequency is observed in celiac disease, psoriasis, systemic sclerosis, rheumatoid arthritis, and lupus erythematosus patients when compared with healthy controls.<sup>16–20</sup>

Despite their inferred role, we still lack knowledge of the structural and functional details of 3'RRs. In vivo experiments, like fluorescence in situ hybridization and chromosome conformation capture, confirmed a large conformational rearrangement of the DNA, although these methods did not provide specific structural information.<sup>4,21</sup> A first insight into the three-dimensional (3D) structure of the 3'RR was given by previous in silico studies that showed the presence of an inverted duplication. As a palindromic sequence, the inverted duplication may give rise to a hairpin-loop shape, in which the hs1.2 enhancer is located at the apical loop.<sup>8,22,23</sup> The inverted duplication can be recognized in all mammals studied so far, although it does not show similarity among species belonging to different orders, supporting the hypothesis of a function

linked to specific 3D structures.<sup>8,24</sup> The hypothesized 3D structure locates the hs1.2 enhancer in the middle of a loop where the DNA is supposed to assume single-stranded conformation. The human version of this enhancer has allelic variants with an internal Variable Number Tandem Repeat (VNTR), consisting of a 40 nucleotides element (40-mer) repeated from one to four times. Studies aimed at defining a list of the human genetic variations using whole genome and multigenome approaches seem to fail to highlight the peculiarity of these VNTR until recently.<sup>25</sup> The 40-mer contains a binding site for NF- $\kappa$ B and SP1, as demonstrated by electrophoretic mobility shift assays.<sup>14,19,26</sup> This finding suggests a relevant biological role of the enhancer.

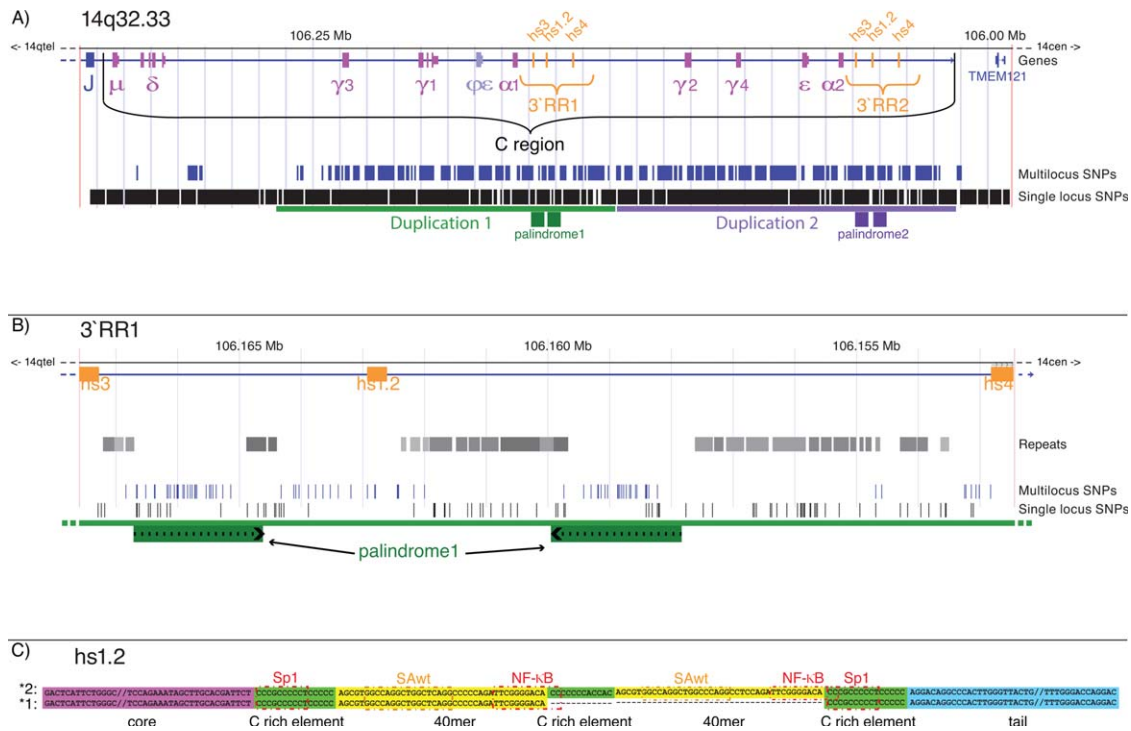
In this work, we analyze in detail a 17-nucleotides DNA sequence (17-mer) spanning a portion of the enhancer hs1.2 that is predicted to form a quadruplex. The 17-mer maps just upstream to the NF- $\kappa$ B and SP1 binding sites and, as a potential quadruplex, can have a relevant regulative role on adjacent transcription factors. Genome wide analyses indeed reported both the proximity of putative quadruplex sequence to transcription factor consensus sites<sup>27</sup> and the association between polymorphisms on quadruplex locus and gene expression.<sup>28,29</sup> Using our strategy, that combines sequence and spectroscopic analysis, we confirmed the ability of this DNA sequence to form in vitro a quadruplex structure and showed that this feature of the enhancer is preserved from evolutionary divergence. This short sequence is thus a strong candidate to constitute a further fundamental component that finely tunes the hs1.2 activity with protein binding properties similar to many regulatory regions as reported by genomic studies.<sup>30</sup>

## RESULTS

### Analysis of 3'RRs and hs1.2 Main Features, and Schematization of Human Structures

Figure 1 shows a scheme of the human regulative regions (3'RR1 and 3'RR2), located downstream of the IgH  $\alpha$ -gene, ranging from a broad view (Figures 1A and 1B) to details of the nucleotide sequence of the main untranslated feature of the regions, i.e. the enhancer hs1.2 (Figure 1C).

We first considered dbSNP build 137<sup>31</sup> to investigate the possibility of distinguishing the two paralogous copies of the 3'RR in human NGS data which would give us the chance to discern the two loci in 1K Genome Project data. We concluded that it is impossible to distinguish between the two copies because of the high level of similarity between the 3'RR1 and 3'RR2. Figure 1B, indeed, shows SNPs in the enhancer regions,



**FIGURE 1** Human 3'RRs and hs1.2-sequence details. (A) Graphic details of the human Ig constant region (C region), harboring the two copies of the 3' Regulatory Region (3'RR1 and 3'RR2) as part of extended in-tandem duplication involving most of the C region. In the two regulatory regions, the two blocks of both palindrome1 (enlarged in section B) and palindrome2 are the inverted repeated branches of the palindromes cited in the text. Internal to both the 3'RRs are the three enhancers, respectively hs3, hs1.2, and hs4, in orange. The blue arrow shows the transcription direction of the Ig genes. In the two bottom lines, the positions of SNPs from the dbSNP build 137. (B) Portion of the human 3'RR1 in higher detail. In green the two branches of the conserved palindromic sequence, embedding a region including hs1.2 with the potential to fold in a big loop of a hairpin structure. (C) Sequences of hs1.2 allele \*1 and \*2, illustrating VNTR structure, transcription factor consensus sites, and quadruplex putative sites. The main difference between the two isoform is the presence of a different number of the 40mer and C rich elements.

distinguishing between those which map to a single genomic region from those mapping in multiple genomic loci (multilocus), as reported in dbSNP build 137. Our preliminary resequencing trial (data not shown) of 3'RR2 indicate that some of the SNPs labeled as “single” are probably multilocus or misassigned, beside the already conspicuous number of multilocus SNPs indicated in the figure. These findings suggest uncertainty about the SNP locations making it impossible to distinguish if both, either or neither of the 3' RRs is conserved within the human species. A possible explanation is that the alignment of short reads from massive sequencing to a reference genome and the consequent identification of polymorphic loci are error-prone in regions that are duplicated such as the 3'RR, and especially close to the hs1.2 enhancers, because of the high similarity level between its evolutionary-recent paralogous copies.

The hs1.2 enhancer maps to the conserved center of each 3'RR copy (Figure 1B). The core and tail fragments of this functional sequence, respectively, when mammal DNAs are compared with each other.<sup>8</sup> Between core and tail of hs1.2 there is a region that, in Catarrhini's variant, showed multiple alleles due to an embedded VNTR of 40 nucleotides, repeated in human version of the enhancer from one to four times. This 40mer contains binding sites for transcription factors, among which are NF- $\kappa$ B and SP1 (Figure 1C), known as mainly involved in cellular responses to stimuli such as bacterial or viral antigens and to a growing list of cellular stresses.<sup>32</sup> Moreover, the 40-mer contains a putative G-quadruplex site. The multiplication of the 40-mer in the various alleles is thus responsible for the proliferation of sequence of putative transcription-factor-binding and G-quadruplex sites within the enhancer.

### Conservation Among Mammals of G-Quadruplex Predicted in hs1.2

Bioinformatics evidence shows that putative quadruplex regions are often found close to transcription factor binding sites.<sup>30,33</sup> We thus scanned the 3'RRs in search of putative quadruplex sites and we found evidence for it within hs3, hs1.2, and hs4, the 3'RRs enhancers that are rich in transcription factor binding sites. We focused our attention on the G-quadruplex inside hs1.2 because of its multiple copies, a feature observed in most primates.<sup>8</sup> In fact, we noted that the four pairs of guanines putatively forming the quadruplex, are in the VNTR element (Figure 1C), and downstream from a core sequence highly conserved in mammals.<sup>8</sup> Figure 1C reports the 17-mer (GGCCAGGCTGGCCAGG labeled as SAwt) including the four GG-pairs within the hs1.2 enhancer that was predicted to form the G-quadruplex. By comparing sequences available in the GenBank database, we found that the SAwt sequence within the hs1.2 is highly conserved among mammals (Table I), with a strong constraint on the guanine pairs necessary for the formation of the quadruplex structure. Some species show the contemporary loss of one of the "canonical" G-pair, and gain of a new G-pair, so the potential to form quadruplex in the 40-mer is unaltered. This finding confirms a restriction to the variability of these four pairs of nucleotides and, consequently, the putative functional role of this element in the enhancer sequence. It should be noted that mouse is the only species that has lost two of the G-pairs in the 40-mer, so the quadruplex site might actually not be present in mice. The rodents, however, evolved a peculiar regulatory mechanism involving three lineage-specific insulators<sup>9</sup> that could complement the quadruplex absence.

### Native Electrophoresis of the Putative Quadruplex Regions

In order to get insights on the behavior in solution of the SAwt sequence, we performed native polyacrylamide gel electrophoresis (PAGE). We decided to buffer the solutions with potassium as a counter-ion to exploit the known tendency of this ion to stabilize quadruplex formation through its ability to bind between the guanine quartets.<sup>34</sup>

Figures 2A and 2B show the PAGE of SAwt at 5 and 100 mM potassium ion concentration and pH 6.5.

Under both conditions, SAwt shows a smearing of bands of several molecular weights.

Two more prominent bands, migrating slower and faster than the unfolded 20-mer oligonucleotide used as a control, are observed.

Thus, the data indicate the presence of multiple species in solution.

A possible explanation is the presence of multiple cytosines in the SAwt sequence and then the possibility of impaired duplex formation. An example is reported in Scheme 1.

In order to exclude the formation of impaired duplex and to verify that the sequence studied can form a quadruplex structure, two different strategies were undertaken:

- Analysis of the SAwt sequence at low pH, where the formation of double strand is hampered;
- Study the behavior of a similar sequence, in which all the cytosines are replaced by thymines, maintaining the correct spacing between guanines nucleotides. Such a sequence will be defined as SAMut (GGTTTGGTTGGTTTTGG).

Here below, both approaches are evaluated in parallel for all techniques used.

Concerning the natural sequence, SAwt, at pH 4.5 (Figures 2C and 2D) no heterogeneity is observed and a single band is present, migrating faster than the control. This band could be attributed to a quadruplex structure since in the presence of saturating amount of potassium, which is known to induce the formation of quadruplex structures, the intensity of this band increases, suggesting a stabilization of the structure.

As far as the SAMut fragment is concerned, the band observed at pH 6.5 did not give a clear indication. However, the observation that at a pH as low as 4.5, a single band of low mobility is observed suggests the formation of a single structure, different from the one observed for the SAwt fragment since they have different mobility.

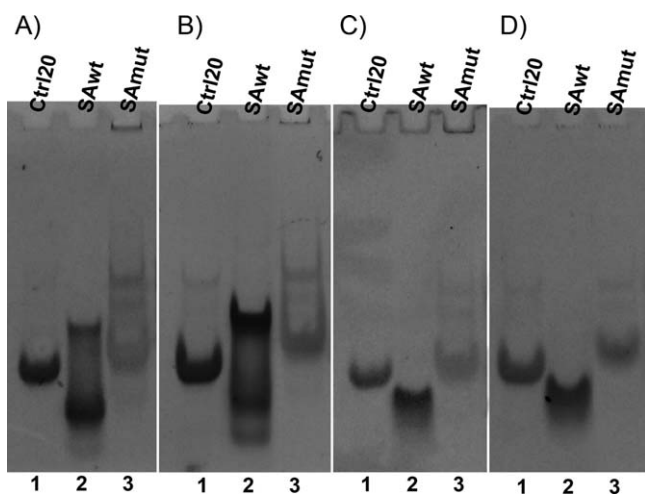
### CD Analysis of the Putative Quadruplex Regions

Figure 3A shows the CD spectra of the native SAwt oligonucleotide at pH 6.5 or 4.5.

At pH 6.5 the spectrum shows two positive peaks around 270 and 285 nm and a negative peak at 240 nm. Addition of saturating amounts of potassium had no effect on the spectrum. It seems difficult to give an unambiguous interpretation of these spectra, but some features resemble the poly(dGdC)-poly(dG)-poly(dC) previously observed with different oligonucleotide DNA sequences.<sup>35,36</sup> At pH 4.5, where the formation of double strands is hampered, the spectrum of SAwt shows a negative peak around 260 nm and strong positive peak at 290 nm that clearly indicate the presence of a quadruplex structure.<sup>37,38</sup> Addition of saturating amounts of potassium has no further effects on the spectrum.

The previous data show that the oligonucleotide at pH 6.5 might form pairings typical of B-DNA, probably deriving from the formation of impaired intermolecular pairings as suggested by the previous electrophoresis data.





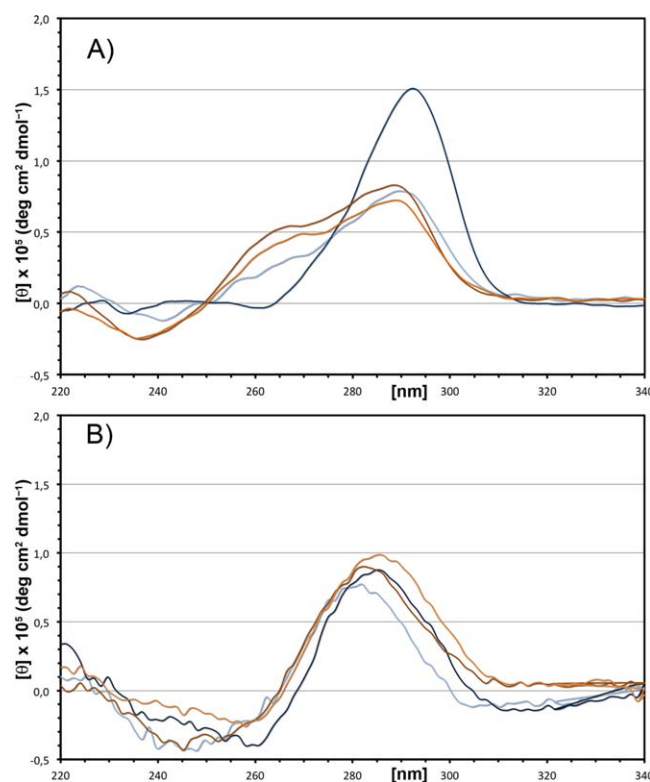
**FIGURE 2** PAGE analysis of the quadruplex region. Native 15% PAGE of SAwt, SAMut and Ctrl20 oligonucleotides in different pH and salt conditions. A) pH 6.5, no KCl; B) pH 6.5, 100 mM KCl; C) pH 4.5, no KCl; D) pH 4.5, 100 mM KCl. In each panel, lane 1 corresponds to a 20-nucleotide oligonucleotide unable to form G-quadruplex, used as a control, while lanes 2 and 3 correspond to SAwt and SAMut, respectively.

This undesirable effect clearly occurs since hundreds of billions of molecules are present *in vitro* but cannot occur *in vivo* where only two strands are present. Thus, we synthesized a different oligonucleotide, SAMut (GGTTTGGTTGGTTTGG), in which all cytosines and adenines have been substituted by thymines but maintaining the right position of guanines. The CD spectrum of this oligonucleotide is shown in Figure 3B. Both at pH 6.5 and at pH 4.5, in the absence or in the presence of saturating amounts of potassium, the spectra show a positive peak around 290 nm indicating the presence of preformed quadruplex structures.

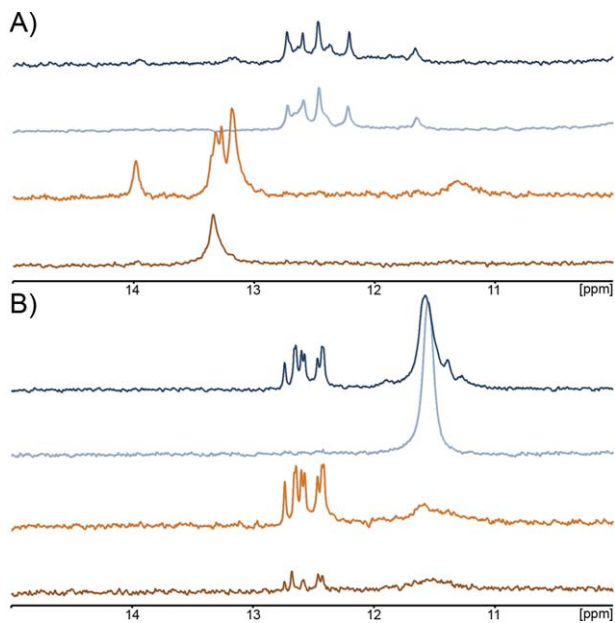
### NMR Analysis of the Putative Quadruplex Regions

Figure 4A shows the NMR spectra collected at 27°C for SAwt at 200  $\mu$ M concentration and different pH values in the absence or in the presence of 100 mM KCl. At pH 6.5, the chemical shifts observed for the imino protons are in 13.00 to 14.00 ppm range, indicative of the formation of double-stranded DNA and/or G•C•G•C tetrads;<sup>39</sup> there is no evidence of quadruplex formation, characterized by a different chemical shift range for imino protons. Since the isolated SAwt oligonucleotide is not autocomplementary but contains several cytosines, we attributed this secondary structure to the formation of impaired alignments (see Scheme 1). This is in agreement with the results obtained by CD and electrophoresis analysis. The number and intensities of peaks in the 13.00 to 14.00 ppm range increase after the addition of 100 mM potassium. Similar spectra were collected at lower concentration of DNA (10  $\mu$ M,

data not shown) to shift the equilibrium toward the single strand in the absence of potassium and possibly toward quadruplex in the presence of potassium but the results were the same, indicating that impaired double strands are very stable. The sample in the presence of a saturating amount of potassium was then kept at 95°C for 10 min and then slowly allowed to anneal but the final secondary structure observed still suggests the presence of double-stranded DNA (data not shown). At lower pH, where the formation of double strands is hampered, the SAwt oligonucleotide shows imino protons peaks in the range 12.00 to 13.0 ppm that are clearly indicative of quadruplex secondary structure, even in the presence of low amounts of potassium. Since potassium is present in the



**FIGURE 3** CD spectra of the quadruplex region. CD spectra of (A) the wild-type 17-mer DNA oligonucleotide used in this study (SAwt) and (B) the mutated 17-mer DNA oligonucleotide in which all nonguanine nucleotides are substituted with thymines (SAMut). The spectra are recorded at different pH and potassium concentration. The spectra at pH 6.5 are the brown (5 mM potassium phosphate buffer) and the orange line (5 mM potassium phosphate buffer, 100 mM KCl). The spectra at pH 4.5 are the light blue (5 mM potassium acetate buffer) and the dark blue line (5 mM potassium acetate buffer, 100 mM KCl). In (A) the two positive peaks at 260 nm and 290 nm are indicative of G•C double strand formation while the small negative and positive peaks at 260 nm and 290 nm are indicative of unimolecular quadruplex. In (B) the negative and positive peaks at 260 nm and 290 nm are indicative of quadruplex formation at pH 4.5.



**FIGURE 4** NMR spectra of the quadruplex region.  $^1\text{H}$  NMR spectra of (A) SAWt and (B) SAMut recorded at different pH and potassium concentrations. Brown spectra refers to pH 6.5, 5 mM potassium phosphate buffer; orange spectra refers to pH 6.5, 5 mM potassium phosphate buffer, 100 mM KCl; light blue spectra refers to pH 4.5, 5 mM potassium acetate buffer; dark blue spectra refers to pH 4.5, 5 mM potassium acetate buffer, 100 mM KCl. Peaks from 13.0 to 14.0 ppm are indicative of double strand formation; narrow peaks between 11.5 and 13.0 ppm are indicative of quadruplex formation. The broad peak around 11.5 ppm belongs to unstructured DNA.

potassium acetate buffer, the formation of quadruplex structure appears already at minimal amounts of potassium. This is in agreement with the CD data that suggest a quadruplex structure under these conditions (Figure 3). Increasing the potassium concentration induces small but detectable changes in the NMR spectrum with small shift of some resonances and different relative intensity of some signals. This could indicate a slight rearrangement of the structure that could well explain the differences observed in the CD spectra. Furthermore, low intensity signals are observed in the same position of the signals observed at pH 6.5. Such signals could indicate a minor form in equilibrium with the quadruplex.

To verify that the double-stranded structure observed at pH 6.5 arises from impaired GC alignment and “Watson-Crick” bonds we also collected NMR spectra of SAMut, the mutant oligonucleotide in which all adenines and cytosines were replaced by thymines. The position of guanines is maintained in this oligonucleotide but impaired alignments cannot be formed. The spectra of this oligonucleotide at different pH values are reported in Figure 4B and show that the imino protons peaks are in the range 11.5 to 12.5 ppm in the presence of

potassium both at neutral and at acidic conditions. This finding indicates the absence of B-DNA formation and the formation of a typical quadruplex DNA in all conditions. The broad peak around 11.5 ppm is indicative of residual, unfolded, DNA.

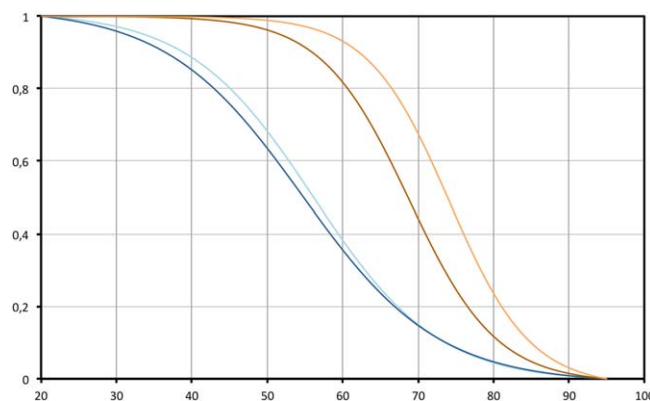
These results clearly indicate the formation of quadruplex structures for both DNA oligonucleotides.

### Thermal Denaturation of the Putative Quadruplex Region

To gain further insight we analyzed the melting profiles of the SAWt fragment present in the human genome. Figure 5 shows the melting curves of SAWt. At pH 4.5 the calculated  $T_m$  at 5 mM and 100 mM potassium are about 55.5°C and 54.8°C, respectively. These values are compatible with the formation of a unimolecular quadruplex as suggested by the previous data. At pH 6.5 the  $T_m$  in the presence of 5 and 100 mM potassium are 68.5°C and 73.8°C, respectively. These values are unusually high for a 17-mer and may reflect the formation of multi-stranded DNA structures.

### SAwt Oligonucleotide Sequence is Conserved Among Primates and Mammals

We identified six loci in the human genome showing 100% identity with the SAWt sequence of hs1.2 enhancers in the 3'RRs (Table II). One of the six loci maps in a large gene-desert region with unidentified function, a second one is within the translated sequence of DCAF12L1 (a member of the WD repeat protein family), and the four others map in transcribed regulatory regions, just like the two hs1.2 copies (Table II). However, when alignment to primate or mammal genome



**FIGURE 5** Melting analysis of the quadruplex region. CD Melting curves of SAWt collected at different pH and different potassium concentrations. Brown, pH 6.5, 5 mM potassium phosphate buffer; orange, pH 6.5, 5 mM potassium phosphate buffer, 100 mM KCl; light blue, pH 4.5, 5 mM potassium acetate buffer; dark blue, pH 4.5, 5 mM potassium acetate buffer, 100 mM KCl.

**Table II Human Genomic Regions Identical to the SAwt Oligonucleotide, and Conservation of These Putative Quadruplex Loci in Mammal Genomes**

chr	Start	End	Strand	Involved Genes	Chromatin State Segmentation	GG-pairs conservation											
						Primate						Others Mammals					
						G1	G2	G3	G4	G1	G2	G3	G4				
2	129064998	129065014	+	HS6ST1 (intron)	Strong enhancer in epidermal keratinocytes (NHEK)	Y	Y	Y	N	N	—	—	—	—	—		
5	1642717	1642733	+	—	Weak/poised enhancer in embryonic stem cells (H1-hESC)	Y	Y	N	N	N	—	—	—	—	—		
6	155541590	155541606	+	TIAM2 (intron)	Weak/poised enhancer in three cell-lines (HUVEC, HMEC, HSMM)	N	Y	Y	N	N	N	N	N	N	Y		
12	124571796	124571812	+	ZNF664-FAM101A (intron)	Insulator in lung fibroblasts (NHLEF)	Y	Y	Y	Y	Y	Y	Y	N	N	Y		
14	106041694	106041710	—	—	hs1.2 in 3'RR2: strong enhancer in B-lymphocyte, lymphoblastoid (GM12878)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		
14	106162771	106162787	+	—	hs1.2 in 3'RR1: strong enhancer in B-lymphocyte, lymphoblastoid (GM12878)	Y	Y	Y	Y	Y	Y	Y	Y	Y	—		
15	82202427	82202443	—	—	—	Y	Y	Y	N	N	N	N	N	N	N		
X	125686156	125686172	+	DCAF12L1 (exon)	Active promoter in umbilical vein endothelial cells (HUVEC)	N	N	Y	Y	Y	Y	Y	Y	Y	Y		

The SAwt sequence was searched in human genome to find similar putative-quadruplex loci, and then each locus was searched in mammals to detect corresponding orthologous loci. Five out of six human sites not belonging to the chr14 have the four guanines pairs (G1, G2, G3, and G4) not conserved in primates. The finding suggests higher positive pressure acting on the chr14 and chr12 sequences. Data in the Chromatin State Segmentation column are from ENCODE ChIP-seq experiments.<sup>40</sup>



of each of the SAwt-similar loci is considered, some interesting indications can be found. In nonhuman primates, all six sequences show some variation with respect to the SAwt sequence that alter the four GG pairs needed for the quadruplex formation, suggesting relaxed selection in these species at those loci. By contrast, when the 3'RR1 and 3'RR2 loci are considered, the GG pairs are conserved in all primates and in many mammal genomes (Tables I and II), suggesting a strong negative selection against mutations at these positions and in turn a possible conserved functional role of this fragment of the enhancer.

## DISCUSSION

Different DNA conformations modify the availability of consensus sequences for the binding of specific proteins and 3D structures in the binding region which can finely tune the binding process.<sup>5,27,30,41–43</sup> In many cases, the DNA molecule must be in an open and accessible state to allow the binding of specific proteins and the 3D structures cooperate to define the actual condition of the embedded consensus sites. This process has been called chromatin remodeling and requires the activity of proteins necessary to maintain the chromatin in an accessible state for nuclear factors like NF- $\kappa$ B.<sup>41</sup> Specific chromatin modifications and configurations are in fact necessary for NF- $\kappa$ B proteins to bind cognate kB motifs.<sup>28</sup> Thus, the presence of kB sites appears to be a requirement but is not sufficient for gene induction.<sup>40</sup> The effectiveness of the consensus site is influenced by the genomic and nuclear environment in which it is embedded.<sup>42</sup>

We previously observed the possibility of a 3D palindromic structure that could modulate the role of the 3'RR by formation of a hairpin structure, in which the hs1.2 enhancer locates at the middle of the apical loop.<sup>8</sup> This site was probably also tandem-repeated in the Catarrhini-parvorder ancestor, in which also the VNTR of hs1.2 originated.<sup>8</sup> These remarks suggest that the current version of the enhancer may have evolved from an ancestral sequence including the core and a single copy of the quadruplex site.

In this article, we show that a quadruplex structure might form inside the hs1.2 enhancers in the 3'RRs (Figures 2 and 3) and we speculate a possible functional role for it in exposing DNA in order to modulate the binding of transcription factors. We show that the hs1.2 short oligonucleotide required for the quadruplex structure formation is highly conserved in primates and mammals, unlike other identical sequences in the human genome, suggesting that strong negative selection is acting on this region and corroborating the hypothesis of a functional role of the quadruplex 3D structures.

The formation of a quadruplex structure indicates the presence of a newly identified functional region within the hs1.2 enhancer. Quadruplex formation in Igs had been speculated many years ago,<sup>44</sup> and the present work gives some experimental evidence of such structure. This finding allows the formulation of hypotheses on the function of this polymorphic enhancer suggesting a possible modulating effect in the equilibrium among the alternative 3D forms that can modulate the binding of NF- $\kappa$ B and SP1 nuclear factor. The predicted quadruplex sequence in fact lies just upstream of the NF- $\kappa$ B transcription factor consensus site (Figure 1C). In this context it must be noted that both NF- $\kappa$ B and Sp1 bind to double-stranded DNA. Thus, a possible role of the quadruplex structure may be to shift the equilibrium of the 3'RR region towards the hairpin structure, stabilizing this alternative conformation of the DNA. The final effect of this conformational change is the inhibition of transcription factors binding since the DNA is single-stranded. An alternative possibility is the formation of a local 3D structure, the quadruplex, along the double-stranded helix, which creates a steric hindrance to the binding of transcription factors to their consensus sites.

These results support our previous findings on physiologic and pathologic phenotypes that clarified the modulatory function of the hs1.2 alleles in Ig production. Taken together, these studies reveal the relevant role of the 3D structures of DNA for regulation of Igh expression and B cell activity.

## MATERIALS AND METHODS

### In Silico Analyses of hs1.2 Sequence

We searched for copies of the hs1.2 enhancer in three Genbank databases (i.e. genome, htgs, and wgs) by the NCBI Blast algorithm (<http://blast.ncbi.nlm.nih.gov/>).<sup>50</sup> The Genomic Evolutionary Rate Profiling (GERP) score, Single Nucleotide Polymorphisms (SNPs), repeat, and 3'RRs features are analyzed, identified and/or graphically visualized using the UCSC Genome Browser<sup>46</sup> (<http://genome.ucsc.edu/cgi-bin/hgGateway>). The G-quadruplex structure consensus sites were identified by Quadbase<sup>47</sup> (<http://quadbase.igib.res.in>) and QGRS Mapper<sup>48</sup> (<http://bioinformatics.ramapo.edu/QGRS>).

### Spectroscopic Analysis of the DNA Sequence

For the *in vitro* analysis we used two oligonucleotides derived from the polymorphic region of hs1.2 enhancer in the human 3'RRs (Figure 1A; detailed view of the 3'RR1 in Figure 1B): a sequence of 17 nucleotides (17-mer, SAwt: GGCCAGGCTGGCCAGG) selected as a G-quadruplex putative consensus sequence from *in silico* analysis (Figure 1C); the second oligonucleotide was designed to replace all the C and A in the SAwt with T (SAMut: GGTTTGGTTGGTTTGG). Both oligonucleotides were synthesized by Eurofins (Milano, Italy), purified by HPLC and desalted. Samples were denatured at 95°C for 10 min and then slowly cooled before use. Circular dichroism (CD) and

nuclear magnetic resonance (NMR) spectra at pH 6.5 were obtained by dissolving the sample in 5 mM potassium phosphate buffer, while the spectra at pH 4.5 were obtained by dissolving the sample in 5 mM potassium acetate buffer. KCl was added from 3M stock solutions and spectra were collected soon after the addition of potassium. CD spectra were collected on a Jasco J600 Spectrometer (Jasco Inc., Easton, MD) using a 0.1 cm path cell at 300 K. Data were obtained from 200 to 340 nm at 0.2 nm interval, 20 nm/min speed, and averaging over four scans. Samples contained 20  $\mu$ M of DNA. Data are reported as mean residue ellipticity. NMR spectra were collected using concentrations ranging from 10 to 200  $\mu$ M at 400 or 600 MHz on Bruker or Varian spectrometers, respectively. A 10% D<sub>2</sub>O was used for lock and water was suppressed by using a watergate pulse sequence<sup>49</sup> that gives a maximum excitation over the imino proton region. The spectra were collected at different temperatures ranging from 2 to 27°C and analysis was done using the MNova software (Mestrelab Research, Spain).

### Polyacrylamide Gel Electrophoresis (PAGE)

PAGE analysis has been carried out as previously described.<sup>50</sup> The samples (5  $\mu$ M oligonucleotides in 5 mM potassium acetate buffer, pH 4.5 or 6.5) were heated to 95°C for 10 min, and then annealed by slowly cooling to room temperature with or without 100 mM KCl. After incubation at 4°C for at least 1 h, the samples were run on 18% native PAGE (19:1 acrylamide:bisacrylamide ratio) with TAE as electrophoretic running buffer, containing the same salt concentration of the run samples. Each run was performed at room temperature for about 1 h and 30 min at 130 V. The gels were stained with SYBR Safe DNA gel stain (Invitrogen).

### CD Melting Analysis

CD spectra were recorded using a JASCO J-715 spectropolarimeter equipped with a Peltier temperature controller, using a quartz cell of 1.0 cm optical path length, in the range 20 to 95°C, increasing the temperature with a 1°C increasing step per minute. The CD melting profiles were obtained plotting the molar ellipticity corresponding to the maximum at about 290 nm as a function of temperature. The resulting curves were normalized obtaining a plot of the folded fraction,  $\alpha$ , versus temperature. The melting temperature ( $T_m$ ) was then deduced as the temperature corresponding to  $\alpha$  equal to 0.5.

### ACKNOWLEDGMENTS

Authors thank Dr. Mariarosaria Tortora for the help with Circular Dichroism measurements. Prof. Mergny of the Université de Bordeaux is gratefully acknowledged for useful discussions during the early stage of the work. Authors are indebted to Annalisa Pastore, MRC London, and to Janez Plavec, University of Ljubljana, for fruitful discussions.

### REFERENCES

1. Kang, J.; Xu, B.; Yao, Y.; Lin, W.; Hennessy, C.; Fraser, P.; Feng, J. *PLoS Comput Biol* 2011, 7, e1002094.

2. Gerstein, M. B.; Kundaje, A.; Hariharan, M.; Landt, S. G.; Yan, K. K.; Cheng, C.; Mu, X. J.; Khurana, E.; Rozowsky, J.; Alexander, R.; Min, R.; Alves, P.; Abyzov, A.; Addleman, N.; Bhardwaj, N.; Boyle, A. P.; Cayting, P.; Charos, A.; Chen, D. Z.; Cheng, Y.; Clarke, D.; Eastman, C.; Euskirchen, G.; Fritze, S.; Fu, Y.; Gertz, J.; Grubert, F.; Harmanci, A.; Jain, P.; Kasowski, M.; Lacroute, P.; Leng, J.; Lian, J.; Monahan, H.; O'Geen, H.; Ouyang, Z.; Partridge, E. C.; Patacsil, D.; Pauli, F.; Raha, D.; Ramirez, L.; Reddy, T. E.; Reed, B.; Shi, M.; Slifer, T.; Wang, J.; Wu, L.; Yang, X.; Yip, K. Y.; Zilberman-Schapira, G.; Batzoglou, S.; Sidow, A.; Farnham, P. J.; Myers, R. M.; Weissman, S. M.; Snyder, M. *Nature* 2012, 489, 91–100.
3. Montavon, T.; Duboule, D. *Trends Cell Biol* 2012, 22, 347–354.
4. Schwab, K. R.; Patel, S. R.; Dressler, G. R. *Mol Cell Biol* 2011, 31, 1503–1511.
5. Hubner, M. R.; Eckersley-Maslin, M. A.; Spector, D. L. *Curr Opin Genet Dev* 2013, 23, 89–95.
6. Birshstein, B. K. *Front Immunol* 2014, 5, 163.
7. Garot, A.; Marquet, M.; Saintamand, A.; Bender, S.; Le Noir, S.; Rouaud, P.; Carrion, C.; Oruc, Z.; Bebin, A. G.; Moreau, J.; Lebrigand, K.; Denizot, Y.; Alt, F. W.; Cogne, M.; Pinaud, E. *Proc Natl Acad Sci USA* 2016, 113, 1618–1623.
8. D'Addabbo, P.; Scascitelli, M.; Giambra, V.; Rocchi, M.; Frezza, D. *BMC Evol Biol* 2011, 11, 71.
9. Sepulveda, M. A.; Garrett, F. E.; Price-Whelan, A.; Birshstein, B. K. *Mol Immunol* 2005, 42, 605–615.
10. Ren, X.; Siegel, R.; Kim, U.; Roeder, R. G. *Mol Cell* 2011, 42, 342–355.
11. Park, S. R.; Kim, P. H.; Lee, K. S.; Lee, S. H.; Seo, G. Y.; Yoo, Y. C.; Lee, J.; Casali, P. *Cytokine* 2013, 61, 608–613.
12. Dunnick, W. A.; Shi, J.; Zerbato, J. M.; Fontaine, C. A.; Collins, J. T. *J Immunol* 2011, 187, 4733–4743.
13. Ju, Z.; Chatterjee, S.; Birshstein, B. K. *Mol Immunol* 2011, 49, 297–303.
14. Frezza, D.; Giambra, V.; Mattioli, C.; Piccoli, K.; Massoud, R.; Siracusano, A.; Di Giannantonio, M.; Birshstein, B. K.; Rubino, I. A. *Int J Immunopathol Pharmacol* 2009, 22, 115–123.
15. Giambra, V.; Cianci, R.; Lolli, S.; Mattioli, C.; Tampella, G.; Cattalini, M.; Kilic, S. S.; Pandolfi, F.; Plebani, A.; Frezza, D. *J Immunol* 2009, 183, 8280–8285.
16. Toluoso, B.; Frezza, D.; Mattioli, C.; Fedele, A. L.; Bosello, S.; Faustini, F.; Peluso, G.; Giambra, V.; Pietrapertosa, D.; Morelli, A.; Gremese, E.; De Santis, M.; Ferraccioli, G. F. *Ann Rheum Dis* 2009, 68, 416–419.
17. Frezza, D.; Giambra, V.; Cianci, R.; Fruscalzo, A.; Giufre, M.; Cammarota, G.; Martinez-Labarga, C.; Rickards, O.; Scibilia, G.; Sferlazzas, C.; Bartolozzi, F.; Starnino, S.; Magazzu, G.; Gasbarrini, G. B.; Pandolfi, F. *Scand J Gastroenterol* 2004, 39, 1083–1087.
18. Frezza, D.; Giambra, V.; Toluoso, B.; De Santis, M.; Bosello, S.; Vettori, S.; Triolo, G.; Valentini, G.; Ferraccioli, G. *Ann Rheum Dis* 2007, 66, 1210–1215.
19. Frezza, D.; Toluoso, B.; Giambra, V.; Gremese, E.; Marchini, M.; Nowik, M.; Serone, E.; D'Addabbo, P.; Mattioli, C.; Canestri, S.; Petricca, L.; D'Antona, G.; Birshstein, B. K.; Scorza, R.; Ferraccioli, G. *Ann Rheum Dis* 2012, 71, 1309–1315.
20. Cianci, R.; Giambra, V.; Mattioli, C.; Esposito, M.; Cammarota, G.; Scibilia, G.; Magazzu, G.; Orlando, A.; Sandri, G.; Bianchi,

- L.; Gasbarrini, G. B.; Pandolfi, F.; Frezza, D. *J Invest Dermatol* 2008, 128, 1920–1924.
21. Guo, C.; Gerasimova, T.; Hao, H.; Ivanova, I.; Chakraborty, T.; Selimyan, R.; Oltz, E. M.; Sen, R. *Cell* 2011, 147, 332–343.
  22. Chauveau, C.; Cogne, M. *Nat Genet* 1996, 14, 15–16.
  23. Saleque, S.; Singh, M.; Little, R. D.; Giannini, S. L.; Michaelson, J. S.; Birshtein, B. K. *J Immunol* 1997, 158, 4780–4787.
  24. Saintamand, A.; Vincent-Fabert, C.; Garot, A.; Rouaud, P.; Oruc, Z.; Magnone, V.; Cogne, M.; Denizot, Y. *Nat Commun* 2016, 7, 10730.
  25. Sudmant, P. H.; Rausch, T.; Gardner, E. J.; Handsaker, R. E.; Abyzov, A.; Huddleston, J.; Zhang, Y.; Ye, K.; Jun, G.; Hsi-Yang Fritz, M.; Konkel, M. K.; Malhotra, A.; Stutz, A. M.; Shi, X.; Paolo Casale, F.; Chen, J.; Hormozdiari, F.; Dayama, G.; Chen, K.; Malig, M.; Chaisson, M. J.; Walter, K.; Meiers, S.; Kashin, S.; Garrison, E.; Auton, A.; Lam, H. Y.; Jasmine Mu, X.; Alkan, C.; Antaki, D.; Bae, T.; Cerveira, E.; Chines, P.; Chong, Z.; Clarke, L.; Dal, E.; Ding, L.; Emery, S.; Fan, X.; Gujral, M.; Kahveci, F.; Kidd, J. M.; Kong, Y.; Lameijer, E. W.; McCarthy, S.; Flicek, P.; Gibbs, R. A.; Marth, G.; Mason, C. E.; Menelaou, A.; Muzny, D. M.; Nelson, B. J.; Noor, A.; Parrish, N. F.; Pendleton, M.; Quitadamo, A.; Raeder, B.; Schadt, E. E.; Romanovitch, M.; Schlattl, A.; Sebra, R.; Shabalina, A. A.; Untergasser, A.; Walker, J. A.; Wang, M.; Yu, F.; Zhang, C.; Zhang, J.; Zheng-Bradley, X.; Zhou, W.; Zichner, T.; Sebati, J.; Batzer, M. A.; McCarroll, S. A.; Genomes Project, C.; Mills, R. E.; Gerstein, M. B.; Bashir, A.; Stegle, O.; Devine, S. E.; Lee, C.; Eichler, E. E.; Korb, J. O. *Nature* 2015, 526, 75–81.
  26. Michaelson, J. S.; Singh, M.; Snapper, C. M.; Sha, W. C.; Baltimore, D.; Birshtein, B. K. *J Immunol* 1996, 156, 2828–2839.
  27. Raiber, E. A.; Kranaster, R.; Lam, E.; Nikan, M.; Balasubramanian, S. *Nucleic Acids Res* 2012, 40, 1499–1508.
  28. Natoli, G. *Cold Spring Harb Perspect Biol* 2009, 1, a000224.
  29. Baral, A.; Kumar, P.; Halder, R.; Mani, P.; Yadav, V. K.; Singh, A.; Das, S. K.; Chowdhury, S. *Nucleic Acids Res* 2012, 40, 3800–3811.
  30. Kumar, P.; Yadav, V. K.; Baral, A.; Saha, D.; Chowdhury, S. *Nucleic Acids Res* 2011, 39, 8005–8016.
  31. Sherry, S. T.; Ward, M. H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E. M.; Sirotkin, K. *Nucleic acids Res* 2001, 29, 308–311.
  32. Chen, L. F.; Greene, W. C. *Nat Rev Mol Cell Biol* 2004, 5, 392–401.
  33. Mani, P.; Yadav, V. K.; Das, S. K.; Chowdhury, S. *PLoS One* 2009, 4, e4399.
  34. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res* 2006, 34, 5402–5415.
  35. Segers-Nolten, G. M.; Sijtsema, N. M.; Otto, C. *Biochemistry* 1997, 36, 13241–13247.
  36. Kypr, J.; Kejnovska, I.; Renciuik, D.; Vorlickova, M. *Nucleic Acids Res* 2009, 37, 1713–1725.
  37. Karsisiotis, A. I.; Hessari, N. M.; Novellino, E.; Spada, G. P.; Randazzo, A.; Webba da Silva, M. *Angew Chem Int Ed Engl* 2011, 50, 10645–10648.
  38. Rujan, I. N.; Meleney, J. C.; Bolton, P. H. *Nucleic Acids Res* 2005, 33, 2022–2031.
  39. Lim, K. W.; Alberti, P.; Guedin, A.; Lacroix, L.; Riou, J. F.; Royle, N. J.; Mergny, J. L.; Phan, A. T. *Nucleic Acids Res* 2009, 37, 6239–6248.
  40. Cheng, C.; Alexander, R.; Min, R.; Leng, J.; Yip, K. Y.; Rozowsky, J.; Yan, K. K.; Dong, X.; Djebali, S.; Ruan, Y.; Davis, C. A.; Carninci, P.; Lassman, T.; Gingeras, T. R.; Guigo, R.; Birney, E.; Weng, Z.; Snyder, M.; Gerstein, M. *Genome Res* 2012, 22, 1658–1667.
  41. Natoli, G. *Immunol Rev* 2012, 246, 183–192.
  42. Dong, X.; Greven, M. C.; Kundaje, A.; Djebali, S.; Brown, J. B.; Cheng, C.; Gingeras, T. R.; Gerstein, M.; Guigo, R.; Birney, E.; Weng, Z. *Genome Biol* 2012, 13, R53.
  43. Maizels, N. *EMBO Rep* 2015, 16, 910–922.
  44. Sen, D.; Gilbert, W. *Nature* 1988, 334, 364–366.
  45. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J Mol Biol* 1990, 215, 403–410.
  46. Meyer, L. R.; Zweig, A. S.; Hinrichs, A. S.; Karolchik, D.; Kuhn, R. M.; Wong, M.; Sloan, C. A.; Rosenbloom, K. R.; Roe, G.; Rhead, B.; Raney, B. J.; Pohl, A.; Malladi, V. S.; Li, C. H.; Lee, B. T.; Learned, K.; Kirkup, V.; Hsu, F.; Heitner, S.; Harte, R. A.; Haeussler, M.; Guruvadoo, L.; Goldman, M.; Giardine, B. M.; Fujita, P. A.; Dreszer, T. R.; Diekhans, M.; Cline, M. S.; Clawson, H.; Barber, G. P.; Haussler, D.; Kent, W. J. *Nucleic Acids Res* 2013, 41, D64–D69.
  47. Rawal, P.; Kummarasetti, V. B.; Ravindran, J.; Kumar, N.; Halder, K.; Sharma, R.; Mukerji, M.; Das, S. K.; Chowdhury, S. *Genome Res* 2006, 16, 644–655.
  48. Kikin, O.; D’Antonio, L.; Bagga, P. S. *Nucleic Acids Res* 2006, 34, W676–W682.
  49. Piotto, M.; Saudek, V.; Sklenar, V. *J Biomol NMR* 1992, 2, 661–665.
  50. Micheli, E.; Martufi, M.; Cacchione, S.; De Santis, P.; Savino, M. *Biophys Chem* 2010, 153, 43–53.

*Reviewing Editor: Kenneth Breslauer*