Possible traces of resonance signaling in the genome

Ivan Savelyev [1], Max Myakishev-Rempel [1,2,3]

1: Localized Therapeutics, San Diego, CA, USA

2: DNA Resonance Lab, San Diego, CA, USA

3: Transposon LLC, San Diego, CA, USA

ABSTRACT

Although theories regarding the role of sequence-specific DNA resonance in biology have abounded for over 40 years, the published evidence for it is lacking. Here, the authors reasoned that for sustained resonance signaling, the number of oscillating DNA sequences per genome should be exceptionally high and that, therefore, genomic repeats of various sizes are good candidates for serving as resonators. Moreover, it was suggested that for the two DNA sequences to resonate, they do not necessarily have to be identical. Therefore, the existence of sequences differing in the primary sequence but having similar resonating sub-structures was proposed. It was hypothesized that such sequences, named HIDERs, would be enriched in the genomes of multicellular species. Specifically, it was hypothesized that delocalized electron clouds of purine-pyrimidine sequences could serve as the basis of HIDERs. The consequent genomic analysis confirmed the enrichment of purine-pyrimidine HIDERs in a few selected genomes of mammals, an insect, and a plant, compared to randomized sequence controls. Similarly, it was suggested that hypothetical delocalized proton clouds of the hydrogen bonds of multiple stacked bases could serve as sequence-dependent hydrogen-bond-based HIDERs. Similarly, the enrichment of such HIDERs was observed. It is suggested that these enrichments are the first evidence in support of sequence-specific resonance signaling in the genome.

Ninety-seven years ago, Alexander Gurwitsch proposed the existence of a morphogenetic field that is created by the body and is responsible for developing and maintaining the shape of the body (Gurwitsch, 1922). He and others demonstrated that biological organisms influence the development of each other at short distances and that some of this influence is blocked by optical filters, suggesting that the morphogenic field is of an electromagnetic nature (Gurwitsch, 1988; Volodyaev and Beloussov, 2015). In 1968, Frohlich predicted that in the presence of constant energy flux, cell and organelle membranes produce coherent waves in the millimeter-wave region, thus creating a coherent state and enabling electric wave signaling in living organisms (Frohlich, 1988). In 1973, Miller and Web further proposed that it is DNA that is producing the morphogenic field and that the genomic code is directly sending and receiving the information from the morphogenic field (Miller and Webb, 1973). The experiments verifying the existence of biological fields involve two samples such as cell culture aliquots in sealed quartz cuvettes separated by optical filters. When one of the aliquots is perturbed, the second one may catch a signal that is transferred non-chemically and is blocked by light-impermeable filters. Such effects are often referred to as "non-chemical cell-cell communication" and are reviewed in refs (Cifra et al., 2011; Scholkmann et al., 2013; Trushin, 2004; Xu et al., 2017). Burlakov experimentally demonstrated that the optical distortion by quartz retroreflectors of the field produced by fish embryos causes developmental abnormalities, thus confirming that the field is morphogenic and electromagnetic (Burkov et al., 2008; Burlakov et al., 2012).

Although the existence of the field and its morphogenic and electromagnetic nature have been demonstrated, the involvement of DNA in its generation, proposed in 1973 by Muller and Webb, remains unproven. Many models for oscillations in DNA have been proposed that involve the movement of groups of atoms in DNA

(referred to here as mechanical oscillations) (Scott, 1985; Volkov and Kosevich, 1987). The spectroscopic detection of coherent mechanical oscillations in DNA was reported to be in the THz range (Sajadi et al., 2011). We proposed that in addition to mechanical oscillations in DNA, the base stack displays oscillations of delocalized electron clouds (Polesskaya et al., 2018), as well as of delocalized proton clouds of the hydrogen bonds (Savelyev et al., 2019). Moreover, we suggested that these oscillations occur in a DNA sequence-dependent manner and provide the primary medium for the formation of the morphogenic field. We also suggested that since electron and proton clouds have low mass and are located inside the base stack, they do not cause significant movement of the heavier DNA atoms and the surrounding water, thereby avoiding the thermal dissipation of energy. We suggested that, therefore, the electron and proton cloud oscillations in the base stack are a more likely medium for the morphogenic field than the heavier atoms of DNA (Polesskaya et al., 2018; Savelyev et al., 2019).

We further proposed that electroacoustic resonances between similar DNA sequences form the basis of signaling within the genome and coordinate the function of the cell. We also suggested possible mechanisms by which these oscillations are channeled by the microtubules from one nucleus to another, forming an oscillation network of the body. This way, we transformed an idea of a diffuse morphogenic field into a model of the morphogenic field traveling between the nuclei via tunnels. This also explains how nature may avoid the dissipation of the electroacoustic signals in tissues (Savelyev et al., 2019). We further implicated genomic repeats as primary candidate sequences to serve as resonators. We suggested that since the 300 base pair-long Alu repeat occurs 1.1 million times in each of our cells, it is the ideal candidate for serving as a resonator by the mere number of copies improving the sustainability of oscillations and reducing the dissipation of the signal. We also suggested that the primary function of genomic repeats such as telomeric, centromeric, simple repeats, and transposable elements is to support the resonance signaling in the genomes of complex organisms. Furthermore, we proposed (Savelyev et al., 2019) that this resonance signaling system is deliberately supported by the cells via the flux of ATP and other biochemical energy, in accordance with the Frohlich models (Fröhlich, 1968). We suggested that similar DNA sequences resonate with each other, forming a resonating network within the nucleus, between the nuclei and across the organism. During this process, some of the repetitive sequences may be energized by chemical processes, and their oscillations may be transmitted along the base stack, causing oscillations in similar sequences. This way, conformational changes in the chromatin in one location can lead to conformational changes in the chromatin of similar DNA sequences, allowing for resonance signaling within the nucleus, and across the organism. We suggested that this process is deliberate, developed by evolution for higher organisms, and that the cell spends ATP and other types of chemical energy on supporting this resonance genomic signaling. This way, the chromatin immediately and mechanistically mediates the interaction between the electromagnetic resonance signaling and molecular signaling in a DNA sequence-specific manner. During this signaling process, the resonance properties of the DNA sequences provide specificity, while the ATP energy allows for the amplification of electromagnetic resonance signals and their conversion to molecular signals. For example, oscillations in some Alu sequences may be induced by ATP-dependent chromatin remodeling factors, and these oscillations may be transmitted via the base stack to the second group of Alu elements. Via electromagnetic resonance, the Alu elements of the second group begin resonance oscillation, which is amplified by the ATP-dependent chromatin remodeling factors bound to them, causing chromatin opening and transcription. This proposed mechanism would explain why Alu elements are enriched in the gene promoters (Savelyev et al., 2019).

Although there are experimental demonstrations of morphogenic field effects, to our knowledge, the involvement of DNA in its formation is yet to be proven. The prediction of oscillation frequencies in DNA is not trivial since it is likely that DNA could support several oscillation modes, including those of mechanical, electron and proton clouds. Since DNA is wrapped around nucleosomes, the chromatin state should also be considered. We suggest that sequence-specific oscillations in DNA can spread over a wide range of frequencies.

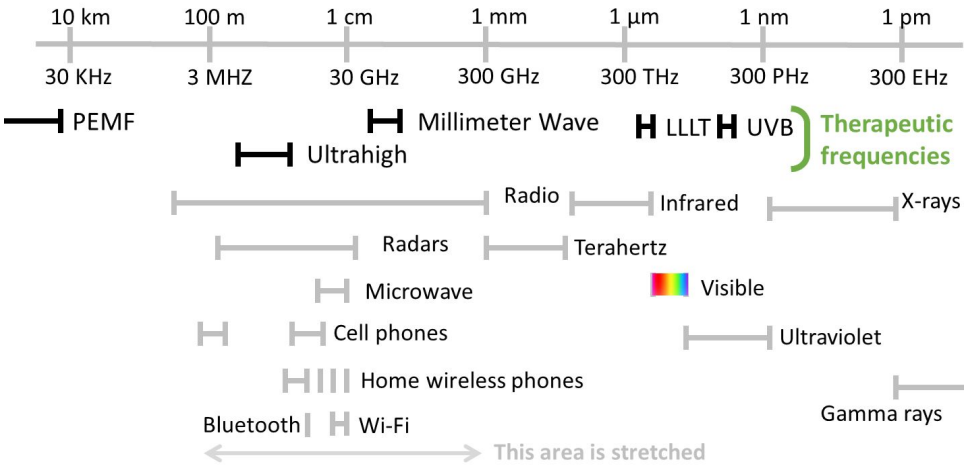
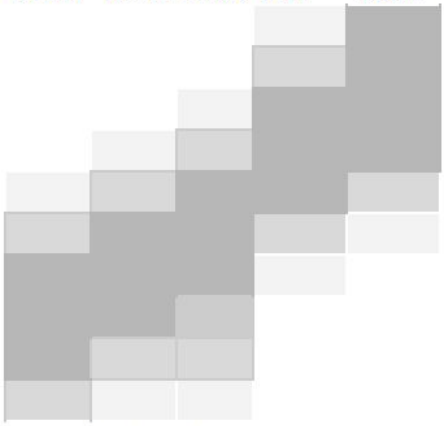


Fig. [Spectrum] Frequency ranges used for therapy. (LLLT – low-level light therapy, PEMF - pulsed electromagnetic field).

Some insight might be obtained from electromagnetic frequencies used in physical therapy. Especially informative would be those frequencies, which produce effects at extremely low power, suggesting that they tap onto electromagnetic resonance signaling. Such frequencies are shown in Fig. [Spectrum]. Specifically, the following therapeutic ranges of electromagnetic frequencies exhibit significant effects at low power, and thus are likely to be tapping existing signaling pathways: pulsed electromagnetic field therapy (Binder et al., 1984), ultra high-frequency therapy (Lushnikov et al., 2004), millimeter wave therapy (Usichenko et al., 2003), low-level light therapy (Bjordal et al., 2003), and UVB (Lowe et al., 1991). We suggest that these frequencies are good candidates for resonance oscillations in DNA. Since the frequency depends on the mass of the oscillator, shorter DNA repeats should oscillate at higher frequencies than the longer ones. Based primarily on these assumptions, we propose the following approximate prediction of resonance frequencies of the genomic repeats, **Table [Wavelengths]**. Note that the natural wavelength of the oscillator can be much larger than its size. Recently, radio communication saw the development of nanomechanical magnetoelectric (ME) antennas, which resonate at wavelengths 1000 times larger than their size (Nan et al., 2017; Shi et al., 2016). An additional conversion factor is that electromagnetic oscillations are coupled with the acoustic oscillations in biological tissues. Therefore, an electromagnetic wavelength from a therapeutics device may be converted to an acoustic wave in the tissue, thus shortening the wavelength approximately 200,000 times. Although the predictions in Table [Wavelengths] are preliminary and requires experimental testing, they aim to illustrate the possible mechanistic connection between electromagnetic therapies and the proposed resonance genomic signaling.

Table [Wavelengths]: A very approximate prediction of resonance wavelengths of genomic repeats

Repeat unit length			Periodic	Type	wavelength	PEMF	UHF	MWT	LLLT	UVB
					light	37km	0.3m	7mm	800nm	300nm
					sound	186m	1.5um	30nm	4nm	1.5nm
1	bp	0.3 nm	y	simple						
2	bp	0.7 nm	y	simple						
3	bp	1.0 nm	y	simple						
4	bp	1.3 nm	y	simple						
6	bp	2.0 nm	y	telomeric						
171	bp	57 nm	y	centromeric						
260	bp	86 nm	n	MIR						
300	bp	100 nm	n	Alu						
1000	bp	332 nm	n	Mariner						
6000	bp	1992 nm	n	LINE1						



(UHF - ultra high frequency, MWT - millimeter wave therapy)

Since so far, to our knowledge, there has been no published evidence for resonance genomic signaling, we attempted to prove it computationally. Since we believe that the majority of repetitive sequences in the genome are involved in meaningful, resonance signaling, we hypothesized that some of the unique (non-repetitive) sequences in the genome might have evolved to resonate with the genomic repeats. Accordingly, we hypothesized that it is not necessary for the unique sequence to be identical to the repeat, that for resonance, it might need to be only partially similar to the sequence of the repeat: for example, it is possible that some oscillations involve primarily the electron clouds of the aromatic rings (Savelyev et al., 2019). Therefore, only the purine-pyrimidine structure of the resonating sequences should be similar, while their primary sequences can be different. This simplification of the sequence from the primary sequence to the purine-pyrimidine sequence, is hereafter referred to as the "Purine code." Similarly, for the oscillations primarily involving the hypothetical clouds of the delocalized protons of the hydrogen bonds in base pairs, only the patterns of these bonds should be similar, while the primary sequence can be different. This simplification of the sequence from primary to strong/weak (three bonds/two bonds per base pair), is hereafter referred to as the "Strong code." The recoding rules used here are listed in **Fig. [Codes]**.

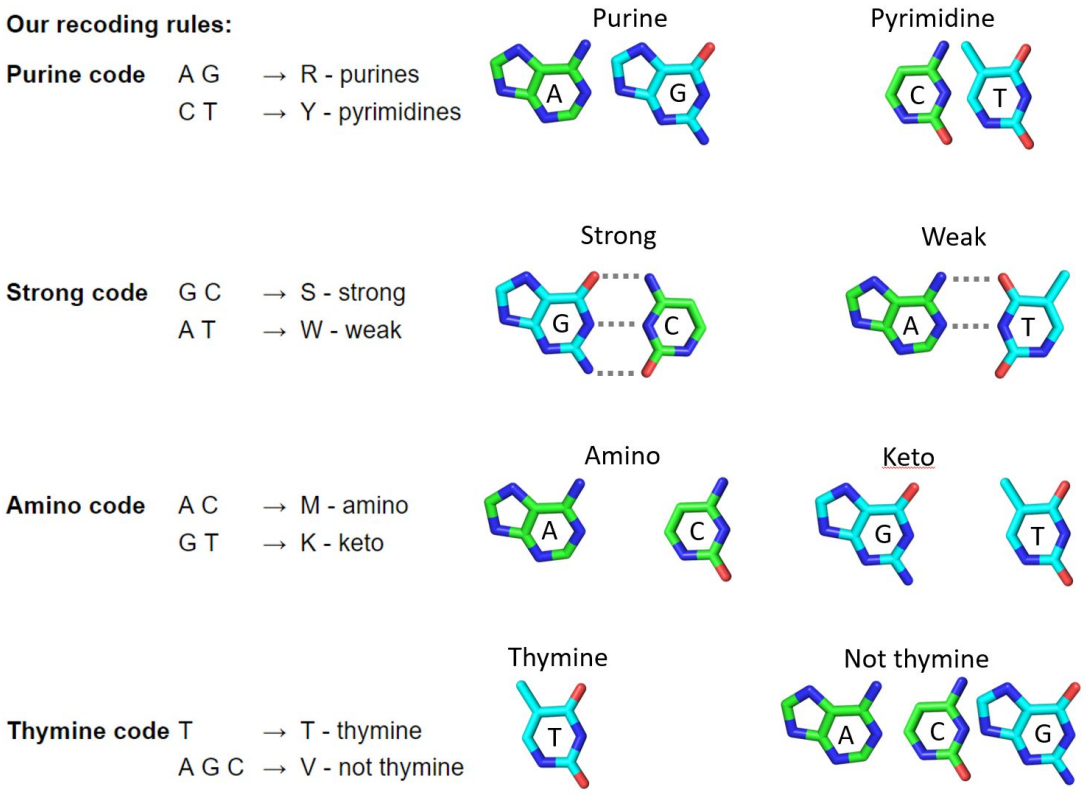


Fig. [Codes]. Recoding schemes used.

Similarly, Amino and Thymine codes were used in the analysis. Therefore, we attempted to search for sequences that are unique (non-repetitive), but become similar to genomic repeats or each other after recoding (simplification). We will refer to them as HIDERs (Homologous If Decoded Elements, Repetitive). In accordance with the four recoding schemes, Table [Codes], four types of HIDERs were analyzed: Purine, Strong, Amine, and Thymine. On the primary sequence level, HIDERs are unique (non-repetitive) sequences, which are identical to each other after recoding. On the physical level, we expect these to be engaged in resonance signaling, and therefore, enriched in the genomes of complex organisms. One of the advantages of such a computational genomics approach is that it is agnostic to the exact physical mechanism of the resonance, allowing verification of its existence prior to the discovery of the mechanism. Once HIDERs are found, their chemical structure may provide an insight into the modes of their resonance.

Methods

The repeats were masked using RepeatMasker (http://repeatmasker.org/) followed by a heuristic removal of repeats using Ugene 1.32.0 (http://ugene.net/). Recoding was done as described in **Fig. [Codes]**. HIDERs were detected by searching for similar pairs of fragments in the recoded sequences using Ugene, analyzed in Google Sheets, and plotted with GraphPad Prism. Randomized sequences were used as controls, see Supplement. To retain the distribution of nucleotide densities along the sequence, randomization was done only on the unmasked parts of the sequence within each 20-nucleotide bin. The significance of enrichment was determined using the t-test.

Results

HIDERs are enriched in genomes compared to randomized controls.

We selected five species for analysis. In addition to humans, we chose mice, drosophila, and arabidopsis as typical model species and the dolphin as a highly developed aquatic mammal.

In each genome, four 90 kb pieces were selected at random, and the repeats were masked. Randomized reference sequences (RAND) were created from the original sequences (ORIG). ORIG and RAND sequences were recoded, as presented in Table [Codes]. In each sequence, pairs of identical strings (HIDERs) longer than 19 bases were identified. The counts of HIDER pairs are shown in Fig. [Counts].

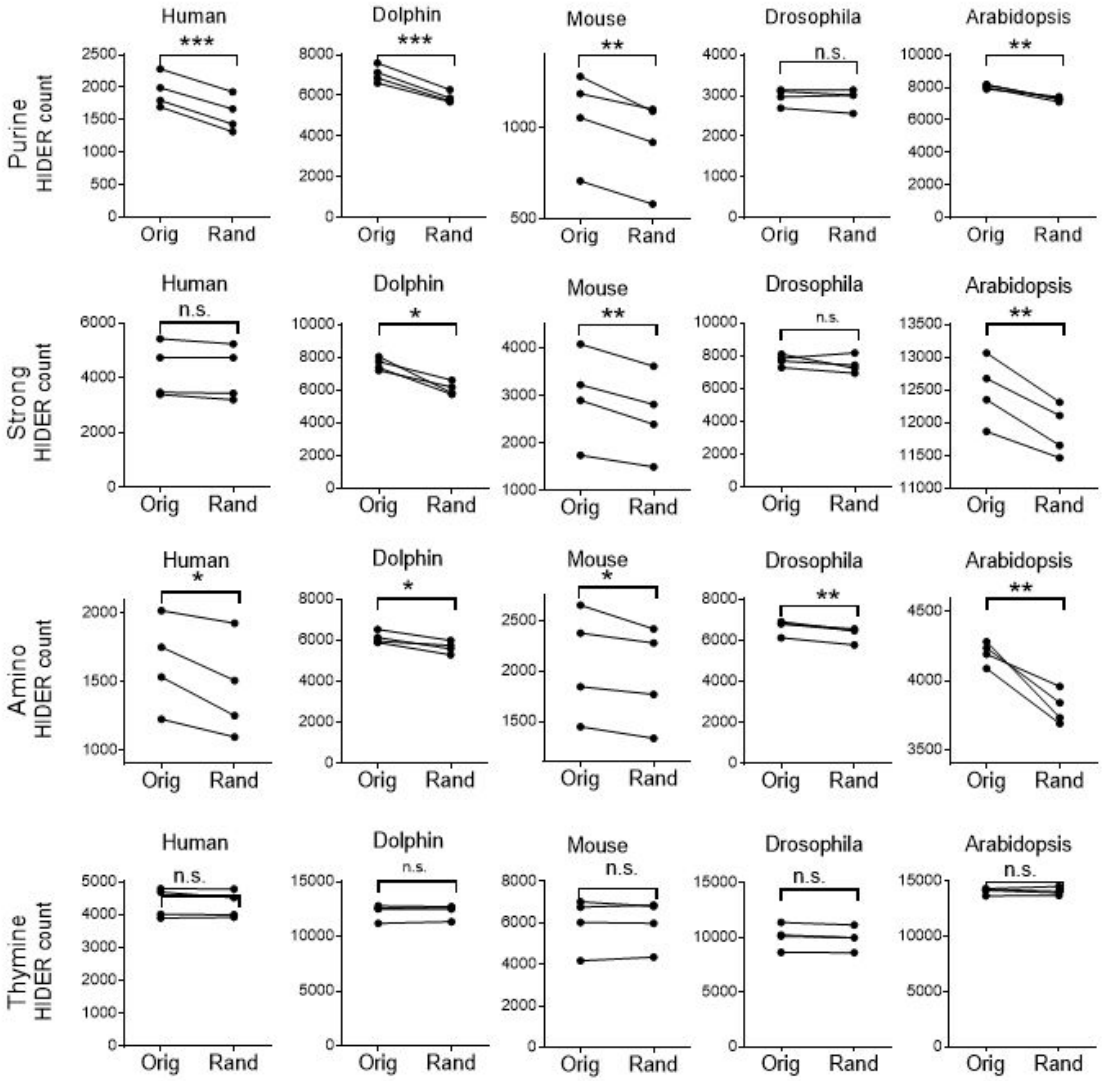


Fig. [Counts] The effect of sequence randomization on the HIDER counts.

The % enrichment of HIDER counts in the original sequences over randomized sequences from the same data is shown in Fig. [Enrichment]

$$\text{Enrichment} = \frac{\text{Count orig.} - \text{Count rand.}}{\text{Count orig.}} \%$$

Code	Human	Dolphin	Mouse	Drosophila	Arabidopsis
Purine	19% ***	16% ***	14% **	4% n.s.	10% **
Strong	3% n.s.	20% *	14% **	4% n.s.	5% **
Amino	12% *	8% *	6% *	5% **	9% **
Thymine	1% n.s.	0% n.s.	0% n.s.	2% n.s.	0% n.s.

Fig. [Enrichment]. Enrichment of HIDER counts in the original sequences over randomized sequences. (* - P < 0.05, ** - P<0.01, *** - P< 0.001, n.s. - nonsignificant)

Among the five tested species, the highest enrichment of HIDERs was found in mammals and the lowest in drosophila. Among the recoding schemes, the highest enrichment was found in the Purine code and the lowest in the Thymine code. The highest statistical significance was observed for the enrichment of Purine HIDERs in humans and dolphins.

Length dependence

In arabidopsis, the Purine HIDERs demonstrated a positive correlation of the HIDERs' enrichment with their length, Fig. [Length]: the enrichment was higher for longer HIDERs.

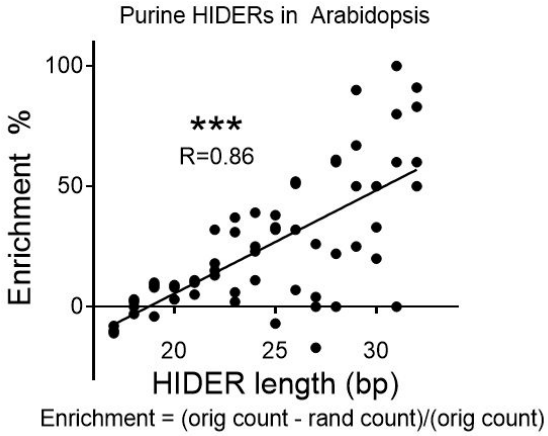


Fig. [Length] Length dependence of Purine HIDER enrichment in arabidopsis.

This suggests that longer HIDERs might be functional and, thus, preferentially selected during the process of evolution. Such correlation was less pronounced in the other species studied.

Discussion

As detailed in the Introduction, our initial motivation was to find sequence-dependent DNA resonators. We realized that for resonance to be sustained, the number of DNA resonators needs to be very high in each cell, in the order of millions of copies. Logically, so-called "junk DNA" made of repetitive elements of various sizes, is the primary candidate for harboring DNA resonators. We suggested that the key resonator in the human genome is the Alu element, which is represented by 1.1 million copies per cell. Then, we proposed that since DNA resonators ought to serve a function in coordinating the operation of the cell and the transfer of information between cells, resonator sequences should evolve to be enriched in the genome. Moreover, we hypothesized that even non-repetitive (unique) sequences might resonate with the repetitive sequences if they support similar modes of oscillation, that is, similar frequencies and patterns of electromagnetic oscillations.

Then, we looked specifically for chemical structures in the DNA, which might support sequence-specific oscillations, and suggested that purine-pyrimidine patterns might be characterized by unique vibrational patterns. Specifically, we hypothesized that the pi-electrons of aromatic rings of multiple nucleobases might form a collective delocalized electron cloud, shape and oscillation pattern, of which would be defined primarily by the purine-pyrimidine sequence. Therefore, DNA sequences, having a different primary sequence but common purine-pyrimidine patterns, might resonate. We called such sequences HIDERs, and suggested that they might be enriched in the genome. Here, we tested this hypothesis and confirmed the enrichment of the HIDERs in the selected mammal species and arabidopsis but not in drosophila, Fig. [Enrichment].

Similarly, we hypothesized that protons of hydrogen bonds of neighboring base pairs would form a delocalized proton cloud (a proton highway). This cloud would be prone to oscillations, and these oscillations would depend on the DNA sequence, specifically on the order in which base pairs with two hydrogen bonds (weak: A, T) and three hydrogen bonds (strong: C, G), respectively, occur in the DNA sequence. As above, we tested whether Strong HIDERs would be enriched in the genome. We observed a significant enrichment in the dolphins, the mice, and arabidopsis, but not in humans or drosophila, Fig. [Enrichment].

Note that both the Purine and Strong codes are simplifying the sequence from four symbols (A, C, G, T) to two symbols (purine/pyrimidine or strong/weak). Some information, including side radicals of the nucleobases, is lost, presumably allowing the HIDERs of different primary sequences to resonate with each other and likely with high-copy genomic repeats. Although we believe that high-copy genomic repeats are the primary resonators in the cell, we focused here on HIDERs since they allow us to test the DNA resonance hypothesis via computational genomics.

To reiterate, to our knowledge, this is the first, although indirect, evidence of DNA resonance in biology. However, although the obtained evidence is encouraging, more research is needed to verify the existence and the mechanistic details of DNA resonance. Computational modeling of the proposed electron and proton clouds of DNA sequences with the use of methods of quantum chemistry and structural biology could verify and substantiate the existence of such sequence-dependent resonating structures. Spectroscopic measurements could substantiate the proposed resonances between various sequences, including the ones highlighted by our analyses.

We are aware that in addition to the DNA resonance explanation, there are possible explanations for the observed enrichments that could involve traditional chemical causes. For example, it is possible that purines are more likely to mutate into each other than into pyrimidines, and vice versa. Therefore, repetitive sequences via the process of mutation might diverge in their primary sequence while retaining their purine-pyrimidine sequence, thus, effectively becoming HIDERs. Similarly, certain repeats might be the targets of transcription factors, which recognize their strong-weak pair sequence while ignoring actual bases. Consequently, certain repeats might diverge in evolution, producing Strong HIDERs. Currently available evolutionary base-substitution rates are not precise enough to enable the delineation of the chemical and resonance causes for the enrichment signal of HIDERs. Therefore, we hope that our results encourage further research, and that the hypothesis of sequence-dependent DNA resonance signaling will be verified more conclusively.

One of the challenges in using computational genomics as a tool for testing hypotheses is the need for the Bonferroni correction in the case of multiple comparisons. To avoid multiple comparisons, we randomly selected the DNA fragments only once, and did not optimize any analysis parameters. The selection and analysis of the data occurred only once. Moreover, to allow for testing the phenomena observed by others, we selected only a few species and only a small part (approximately 3%) of each genome. This way, others could easily reproduce the observed enrichments on untouched data sets.

Acknowledgments.

We thank Oksana Polesskaya, Richard Alan Miller, and James E. Charles for discussion. We thank Martin

Brockman and Juan Moran for programming. The work was funded by Max Myakishev-Rempel.

References

Binder, A., Parr, G., Hazleman, B., Fitton-Jackson, S., 1984. Pulsed electromagnetic field therapy of persistent rotator cuff tendinitis. A double-blind controlled assessment. Lancet 1, 695–698.

Bjordal, J.M., Couppé, C., Chow, R.T., Tunér, J., Ljunggren, E.A., 2003. A systematic review of low level laser therapy with location-specific doses for pain from chronic joint disorders. Aust. J. Physiother. 49, 107–116.

Burkov, V.D., Burlakov, A.B., Perminov, S.V., Kapranov, Y.S., Kufal, G.E., 2008. [Correction of Long Range Interaction Between Biological Objects Using Corner-Cube Reflectors]. Biomedical Radioelectronics 41–48.

Burlakov, A.B., Kapranov, Y.S., Kufal, G.E., Perminov, S.V., 2012. [About possible influence on biological object electromagnetic fields]. Weak and ultraweak fields and radiation in biology and medicine-Proceedings of IV International Congress 111–112.

Cifra, M., Fields, J.Z., Farhadi, A., 2011. Electromagnetic cellular interactions. Prog. Biophys. Mol. Biol. 105, 223–246.

Frohlich, H., 1988. Theoretical physics and biology. Biological Coherence and Response to External Stimuli. Berlin: Springer-Verlag 1–24.

Fröhlich, H., 1968. Long-range coherence and energy storage in biological systems. Int. J. Quantum Chem. 2, 641–649.

Gurwitsch, A., 1922. Über den Begriff des Embryonalen feldes. Wilhelm Roux Arch. Entwickl. Mech. Org. 51, 383–415.

Gurwitsch, A.A., 1988. A historical review of the problem of mitogenetic radiation. Experientia 44, 545–550.

Lowe, N.J., Prystowsky, J.H., Bourget, T., Edelstein, J., Nychay, S., Armstrong, R., 1991. Acitretin plus UVB therapy for psoriasis. Comparisons with placebo plus UVB and acitretin alone. J. Am. Acad. Dermatol. 24, 591–594.

Lushnikov, K.V., Shumilina, Y.V., Yakushina, V.S., Gapeev, A.B., Sadovnikov, V.B., Chemeris, N.K., 2004. Effects of low-intensity ultrahigh frequency electromagnetic radiation on inflammatory processes. Bull. Exp. Biol. Med. 137, 364–366.

Miller, R.A., Webb, B., 1973. Embryonic Holography: An Application of the Holographic Concept of Reality. DNA Decipher Journal 2.

Nan, T., Lin, H., Gao, Y., Matyushov, A., Yu, G., Chen, H., Sun, N., Wei, S., Wang, Z., Li, M., Wang, X., Belkessam, A., Guo, R., Chen, B., Zhou, J., Qian, Z., Hui, Y., Rinaldi, M., McConney, M.E., Howe, B.M., Hu, Z., Jones, J.G., Brown, G.J., Sun, N.X., 2017. Acoustically actuated ultra-compact NEMS magnetoelectric antennas. Nat. Commun. 8, 296.

Polesskaya, O., Guschin, V., Kondratev, N., Garanina, I., Nazarenko, O., Zyryanova, N., Tovmash, A., Mara, A., Shapiro, T., Erdyneeva, E., Others, 2018. On possible role of DNA electrodynamics in chromatin regulation. Prog. Biophys. Mol. Biol. 30, 1e5.

Sajadi, M., Furse, K.E., Zhang, X.-X., Dehmel, L., Kovalenko, S.A., Corcelli, S.A., Ernsting, N.P., 2011. Detection of DNA--Ligand Binding Oscillations by Stokes-Shift Measurements. Angew. Chem. Int. Ed. 50, 9501–9505.

Savelyev, I.V., Zyryanova, N.V., Polesskaya, O.O., Myakishev-Rempel, M., 2019. On The Existence of The DNA Resonance Code and Its Possible Mechanistic Connection to The Neural Code. Neuroquantology 17. https://doi.org/10.14704/nq.2019.17.2.1973

Scholkmann, F., Fels, D., Cifra, M., 2013. Non-chemical and non-contact cell-to-cell communication: a short review. Am. J. Transl. Res. 5, 586–593.

Scott, A.C., 1985. Soliton oscillations in DNA. Phys. Rev. A Gen. Phys. 31, 3518–3519.

Shi, Y., Choi, M., Li, Z., Kim, G., Foo, Z., Kim, H., Wentzloff, D., Blaauw, D., 2016. 26.7 A 10mm3 syringe-implantable near-field radio system on glass substrate, in: 2016 IEEE International Solid-State Circuits Conference (ISSCC). pp. 448–449.

Trushin, M.V., 2004. Distant non-chemical communication in various biological systems. Riv. Biol. 97, 409–442.

Usichenko, T.I., Ivashkivsky, O.I., Gizhko, V.V., 2003. Treatment of rheumatoid arthritis with electromagnetic

millimeter waves applied to acupuncture points--a randomized double blind clinical study. Acupunct. Electrother. Res. 28, 11–18.
Volkov, S.N., Kosevich, A.M., 1987. Conformation oscillations of DNA. Mol. Biol. 21, 797–806.
Volodyaev, I., Beloussov, L.V., 2015. Revisiting the mitogenetic effect of ultra-weak photon emission. Front. Physiol. 6, 241.
Xu, J., Yang, F., Han, D., Xu, S., 2017. Wireless Communication in Biosystems. arXiv [physics.bio-ph].

SUPPLEMENT

Sequences used for the analysis.
Four original and four randomized sequences were investigated,

Random sequence selection method: To avoid the multiple comparison problem in the statistics, the selection of sequences was performed only once. Each sequence was 90Kb long. The selection was achieved using a simple algorithm, and the coordinates were predetermined using a simple rule. The assemblies and the coordinates were as follows:

Human
hg38_dna range = chr1: 100000000-100090000
hg38_dna range = chr1: 100090001-100180000
hg38_dna range = chr1: 100180001-100270000
hg38_dna range = chr1: 100270001-100360000

Dolphin
turTru2_dna range=JH472452:10000-100000
turTru2_dna range=JH472452:181000-271000
turTru2_dna range=JH472452:309250-399250
turTru2_dna range=JH472452:480250-570250

Mouse
mm10_dna range=chr3:32500000-32590000
mm10_dna range=chr3:32590001-32680000
mm10_dna range=chr3:32680001-32770000
mm10_dna range=chr3:32770001-32860000

Drosophila
dm6_dna range=chr2L:200000-290000
dm6_dna range=chr2L:400000-490000
dm6_dna range=chr2L:800000-890000
dm6_dna range=chr2L:1200000-1290000

Arabidopsis
hub_329263_araTha1_dna range=chr3:400000-490000
hub_329263_araTha1_dna range=chr3:600000-690000

hub_329263_araTha1_dna range=chr3:1490000-1580000
hub_329263_araTha1_dna range=chr3:1800000-1890000

Repeat masking

Repeat masking was conducted in two steps. The original sequence was uploaded into the online RepeatMasker service (http://repeatmasker.org/cgi-bin/WEBRepeatMasker), and the repeats were masked with Ns. Then the sequence was masked by the Find Repeats algorithm of the UGENE program (Unipro UGENE http://ugene.net/).

Search for HIDERs in the recoded sequence.

The masked sequence was randomized as described in the Methods section. The original and randomized sequences were transformed into degenerate codes, as shown in Fig. [Codes]. Pairs of identical HIDERs longer than 19 pairs were identified using the Find Repeats algorithm of the UGENE program. The accuracy of the search was tested in part by verifying that the primary source sequences for the pairs of HIDERs were different as intended.