Chapter in: Handbook: "The Cerebellum and Cerebellar Disorders". Ed. Dr. Mario Manto, Springer Verlag. Submitted Oct. 20, Accepted Nov. 1, 2011.

RECURSIVE GENOME FUNCTION OF THE CEREBELLUM: GEOMETRIC UNIFICATION OF NEUROSCIENCE AND GENOMICS

Andras J. Pellionisz[1], Roy Graham[2], Peter A. Pellionisz[3], Jean-Claude Perez[4]

Affiliation

[1] HolGenTech, President, Sunnyvale, California, USA, 94086, holgentech_at_gmail_dot_com
[2] DRC Computer, Vice President, Sunnyvale, California, USA, 94089
[3] UCLA Westwood, California, USA, 90024
[4] IBM Emeritus, Independent researcher, Martignas, France 33127

Abstract

Recursive Fractal Genome Function in the geometric mind frame of Tensor Network Theory (TNT) leads through FractoGene to a mathematical unification of physiological and pathological development of neural structure and function as governed by the genome. The cerebellum serves as the best platform for unification of neuroscience and genomics. The matrix of massively parallel neural nets of fractal Purkinje brain cells explains the sensorimotor, multidimensional non-Euclidean coordination by the cerebellum acting as a space-time metric tensor. In TNT, the recursion of covariant sensory vectors into contravariant motor executions converges into Eigenstates composing the cerebellar metric as a Moore-Penrose Pseudo-Inverse.

The Principle of Recursion is generalized to genomic systems with the realization that the assembly of proteins from nucleic acids as governed by regulation of coding RNA (cRNA) is a contravariant multi-component functor, where in turn the quantum states of resulting protein structures both in intergenic and intronic sequences are measured in a covariant manner by non-coding RNA (ncRNA) arising as a result of proteins binding with ncDNA modulated by transcription factors. Thus, cRNA and ncRNA vectors by their interference constitute a genomic metric. Recursion through massively parallel neural network and genomic systems raises the question if it obeys the Weyl law of Fractal Quantum Eigenstates, or when derailed, pathologically results in aberrant methylation or chromatin modulation; the root cause of cancerous growth. The growth of fractal Purkinje neurons of the cerebellum is governed by the aperiodical discrete quantum system of sequences of DNA bases, codons and motifs. The full genome is fractal; the discrete quantum system of pyknon-like elements follows the Zipf-Mandelbrot Parabolic Fractal Distribution curve.

The Fractal Approach to Recursive Iteration has been used to identify fractal defects causing a cerebellar disease, the Friedreich Spinocerebellar Ataxia – in this case as runs disrupting a fractal regulatory sequence. Massive deployment starts by an open domain collaborative definition of a standard for fractal genome dimension in the embedding spaces of the genome-epigenome-methylome to optimally diagnose cancerous hologenome in the nucleotide, codon or motif-hyperspaces. Recursion is parallelized both by open domain algorithms, and also by proprietary FractoGene algorithms on high performance computing platforms, for genome analytics on accelerated private hybrid clouds with PDA personal interfaces, becoming the mainstay of clinical genomic measures prior and post cancer intervention in hospitals and serve consumers at large as Personal Genome Assistants.

Keywords: Cerebellum, RNA as a Genomic Cerebellum, Covariant functors, Contravariant functors, Recursion, Fractal, Golden ratio, Chaos, Recursive Genome Function, Coordinated Genome Function, Purkinje, Tensor Network Theory, Recursion, FractoGene, Covariant, Contravariant, Metric Tensor, Generalized Coordinates, Sensorimotor Coordination, Cancer, Friedreich Spinocerebellar Ataxia, Fractal Defect, Weyl Law, Zipf-Mandelbrot Parabolic Fractal Distribution, Mandelbrot, Transcription factor, Homeodomain, Homeoprotein, cRNA, ncRNA, cDNA, ncDNA, Junk DNA, Central Dogma, Moore-Penrose

1. Introduction

1.1. Agenda: The Cerebellum as the Platform for the Unification of Neuroscience and Genomics by the Geometric School of Biophysics

Our understanding of both the genome and the brain will remain partial and disjointed until we reach a unification of the intrinsic mathematics of structuro-functional geometry of both – as the first is without question a foundation of the second.

The cerebellum emerged in the past half a Century as the best known neural net of the brain since Moruzzi [1950], Jansen and Brodal [1954], Dow and Moruzzi [1958], Eccles, Ito and Szentágothai [1967]. Thus, this CNS subsystem became a fertile ground of theoretical advances as recently reviewed [Manto, 2008]. It is remarkable that some of the earliest concepts as shown below can be traced back to Centuries, but later they became heavily influenced not only by their underlying philosophies, but also by trendy schools from various periods of history. It was only recently that concepts consolidated into mathematically sophisticated theories of neural networks. For a recent review, see Fiori [2008]. References to Tensor Network Theory (TNT) are too many to list.

The level of mathematical abstraction was challenging, as the dual tensor-representation of covariants and contravariants, while fundamental in mathematics of generalized vector (tensor) calculus; Sylvester [1853], was not well understood in its application for sensory- and motor vectors, in spite of a brilliant encapsulation; Anderson [1990]. Here, embracing the generalization of the concept for covariant "protein signaling" RNA (non-coding; ncRNA) versus contravariant "executory" RNA vectors (coding; cRNA) vectors also calls for cross-disciplinary expertise. While at the introduction of TNT Amari initially went public with a dubious critique; see note added in proof in Pellionisz and Llinás [1985]. However, he reversed face soon; Amari [1991] actively uses covariant and contravariant metric tensors and Riemannian metric tensors as a foundation of "Information Geometry". Yet, mathematical theories of (cerebellar) neural networks had minimal impact on neuroscience in the 20th Century for their mathematics-aversion, prized for US aerospace application; Pellionisz et al [1992] and Germany; Eckmiller [1990]. Decades can be lost if paradigm-shifts are not embraced in a timely manner; Kuhn [1962]. An example is that from the encyclopedic formulation of TNT; Pellionisz [1987] it took two decades for Roy and Llinás [2007] and Fiori [2008] to attempt to improve on it. More interestingly, attempts were aimed at making TNT more *dynamic* and also to extend internal representation of the sensorimotor geometry to the *organization of the self*. Note, that the experimental-theoretical collaboration 34 years earlier started with "*Dynamic* Single Unit Simulation of a Realistic Cerebellar *Network* Model"; Pellionisz and Szentágothai [1973], cited in Pellionisz and Llinás [1985], and the hierarchy of internal representations was also laid out [pp. 268-70].

Genomics of 21st Century might not afford to be as luxurious to let several decades to be wasted. The rapid rise of Genome Sequencing industry must be matched by Genome Analytics; Lander et al [2001], Venter et al [2001], Church [2005], Mardis ER [2006], Gibbs et al. [2007], Collins [2007], Pellionisz [2008b]. After the $3Bn "Human Genome Project" there is a general realization that *"our concepts of genome regulation are frighteningly unsophisticated"*; Venter [2011]. Indeed, instead of "gene regulation" or "genome regulation" a conceptual shift to "multidimensional coordinated genome function" is required.

It is now widely recognized that Genome Informatics simply will not do without massive computing, requiring algorithmic mathematical approaches to program them, and the fact that neural function arises from neural networks that are governed by genomic and epigenomic (hologenomic) factors. As a result, some pioneers of the field of Neural Nets swiftly migrated to become leaders in genome informatics Haussler [1995], and biologists imprinted by the General System Theory; Bertalanffy [1934] a decade ago started to claim that "Genomics became Informatics"; Hood [2002]. A common mathematical underpinning of neuroscience and genomics emerged even before the ENCODE Project led by the US Government concluded in the imperative that *"now the community of scientists have to re-think long-held beliefs"*; Collins [2007]; Pellionisz [2006], Simons and Pellionisz [2006a]. With the hindrance of old dogmas defeated in less than three years, The Principle of Recursive Genome Function rapidly gained ground; Pellionisz [2008a,b], Shapshak et al. [2008], Chiappelli et al. [2008], Cartieri [2009], Pellionisz [2009a,b], Perez [2010], Arneth [2010], Oller [2010], Stagnaro [2011], Stagnaro and Caramel [2011], Elnitski et al. [2011].

With the advances of 21st Century genomics, the cerebellum is not just a neural net for sensorimotor coordination, but lends itself to be a unique platform for unification, how genomic and epigenomic (hologenomic) factors create the physiology as well as pathology of cerebellar organelles (most remarkably, Purkinje neurons), organs (the cerebellum) and organisms (sensorimotor system). While the geometrization of Neuroscience with TNT to arrive at the "*Galilean combination of Simplification, Unification, Mathematization*", Churchland [1986] emerged decades too early, mathematization of Genomics is now an urgent socio-economic necessity. Without advanced mathematics yielding software-enabling algorithms, duties of genomics are impossible to carry out within the narrow boundaries of limited domains. This does not mean, of course, that established disciplines are not to stay, but e.g. as Erez-Lieberman et al. [2009] in their paper amounting to a call by co-author Dr. Lander *"Mr. President, the Genome is Fractal!"*, biochemistry was applied to advance, rather than hinder, a paradigm-shift of the early seminal idea of fractal DNA folding; Grosberg et al. [1988, 1993].

TNT may qualify as the best platform for unification, from neural nets to genomics "top down", and "bottom up" towards consciousness. Beyond establishing encyclopedic use of tensor geometry; Pellionisz [1987], TNT has is experimentally supported for arm-movements by Gielen and Zuylen [1985], Bloedel et al. [1988], Laczkó et al [1988], for gaze-control by Pellionisz [1985], Daunicht and Pellionisz [1987], Pellionisz and Graf [1987], Pellionisz et al. [1991], for vestibulo-collicular sensorimotor system by Laczkó et al [1987], Peterson et al. [1987, 1989], Lestienne et al. [1988]. Belated followership improved upon the pioneering; Roy and Llinás [2007], Fiori [2008].

There is not much question that growth of neural networks, such as those of the cerebellum, are governed by genomic and epigenomic (hologenomic) factors. Likewise, it seems to be beyond reasonable doubt that both genome function and the function (sensorimotor coordination) is deeply rooted in recursion; see the epoch-making concept of "feedback" by Cybernetics; Wiener [1948].

Our understanding of the intrinsic mathematics of both Neuroscience and Genomics has reached the critical mass of mathematical overlap of these two fields of biology. This chapter aims at an algorithmic unification of both neuroscience and genomics by the mathematical means of non-Euclidean tensor and fractal geometry. HoloGenomics unites Neuroscience with Genomics, Epigenomics, in terms of Informatics. Time has come to identify the common geometric roots of genome function and how they govern growth and functioning neuronal networks in both a physiological as well as a pathological manner.

2. Recursion in the Cerebellum

2.1. Review: Philosophies, Theories and Computational Models as Foundations of the School of Cerebellar Recursion

Western philosophies traditionally embraced the age-old "arrow model" of deductive, deterministic timeline and unidirectional "cause and result"; Churchland [1986], DuPré [2008]. This is in contrast to the inductive yin-yang of equilibrium, oscillations and

interdeterminism of Eastern philosophies; Zuangzi [~400 BC]. Theory of Relativity by Einstein and the Principle of Uncertainty by quantum mechanics of Planck-Heisenberg-Schrödinger therefore, shook the intellectual foundations of Western philosophies.

The result was an interesting fork. On one hand "System Theory" was outlined as an attempt to encompass complexity; Bertalanffy [1934]. However, Systems Theory hardly aimed at defining the intrinsic mathematics of living systems. Thus, on the other hand, massive simplifications occurred, to regain temporary balance. Compared to Schrödinger's *"What is Life?"* [1944], too early to know the A, C, T and G quanta of "heredity encoded by covalent bondings on an aperiodical crystal", Cybernetics (Greek, "to govern by feedback"), Wiener [1948] took a "reverse engineering" trend of simplification, almost exclusively based on "feedback". Cybernetics, for its reductionism and relying on concepts of engineering, logical calculus and information theory; McCulloch and Pitts [1943], Shannon [1948] rose with the catapulting digital computing architectures to attain intellectual dominance; Neumann [1958 – see also 2nd Edition with Introduction by Drs. Churchland]. Though Neumann, the inventor of computers warned that the mathematics of computers and brains are profoundly different (the latter remained a mystery with von Neumann's demise) [1958] his tragically short-lived life aborted the break-through to find mathematics intrinsic to neural- (let alone the at that time largely unknown genome) systems. Instead, an even more drastic simplification was dogmatized by Crick [1956, 1970], groundlessly proclaiming the DNA>RNA>PROTEINS to be "an arrow-type open loop". Crick's Dogma oversimplified even the "arrow model" into a single channel of action – clinched by rendering recursion to "Junk DNA" pointless; Ohno [1972] – though "non-coding" DNA is actually 98.7% of the human genome! Crick's Central Dogma and Ohno's Junk DNA obsolete notions were surpassed by The Principle of Recursive Genome Function Pellionisz [2008a,b, 2009a]. Crick said [1970] that if his Central Dogma would be proven to be untrue, it would be necessary to put genomics onto an entirely new intellectual foundation. Now with The Principle of Recursive Genome Function; Pellionisz (2008a) not only it became demonstrably untrue but was superseded by a more advanced theory. The revolution lies in recursion.

The side-track of simplification continued into the overly ambitious notion that science does not need to understand Nature's systems to mimic them. Thus, "Artificial Intelligence" (AI) emerged, see the Perceptron; Minsky and Papert [1969]. AI took off and ruled, in part (by mathematically mistakenly) "proving" that neural nets are incapable of performing the key exclusive "or" operation in mathematical logic. It took pioneers of Neural Nets; Hopfield [1982] to rectify the damage caused by the simplistic course, yet AI was only as recently as in 2003 declared by its originator ineffective.

These trends influenced the expression of the most widely accepted classical concept; Flourens' [1824] that the biological neural networks of the cerebellum function to coordinate sensory and motor information. For the general audience, see Wikipedia and for specialists a recent review; Manto [2008]. The seminal concept was traced back by Finger-Stanley [1994], pp. 211-121 to a quarter of a Millennium to originate implicitly the experiments by Barron Larrey [1760], surgeon to Napoleon; whose experiments *"involved inserting needles into the brains of some pigeons … needles pushed to the cerebellum caused his bird to sway. …, this probably was one of the first experimental demonstrations of the association between cerebellar damage and problems with* coordination*"*. As shown in Finger-Stanley]1994], Di Luigi Rolando [1908] followed-up on Larrey, but largely missed, till Flourens pinned down close to 200 years ago [1824] that *"I have shown that all movements persist after ablation of the cerebellum; they lack only being regulated and* coordinated*"*; pp. 292 and. 212 of Finger-Stanley [1994]. In modern times, Holmes [1939] re-established the concept that *without the cerebellum, coordination is known to be absent* – the syndrome aptly named, even before its *metric tensor* function was discovered as *dysmetria*.

Cerebellar theories were reviewed in Pellionisz [1984], for a recent review see D'Angelo et al [2010]. In the decades of conceptual confusion caused by the conflict of philosophies, the already long established facts of cerebellar coordination slipped into "arrow models" of cerebellar theory, such as describing it as a now known conceptual oversimplification of a "timing device" to set the temporal distance from intention to action, Braitenberg et al [1967]. Perhaps due to the emergence of Minsky's Perceptron as a "learning device", Marr's model [1969] utilized a coincidence for "motor learning". However, he repudiated his concept, since (as he said) motor learning did not explain coordination; Marr [1982. p. 14]. With the untimely decline of his health, a Marr-Albus "learning model" emerged; Albus [1971], see review in Pellionisz [1986]. Featuring the cerebellum as any kind of a "filter device belongs also with the category of "arrow models", since it streamlines the arrow-process of unneeded factors but disregards the cardinal notion of feedback. The cerebellum is conceptually not a timer, not a filter but a transformer, converting the multidimensional vector-expression, from covariant intention tensor to coordinated contravariant execution tensor by means of recursion, see TNT.

2.2. The Concept of Coordinates and their Recursion as Basics of Tensor Network Theory of Cerebellar Neural Nets

In a mathematical sense, as reviewed earlier, Pellionisz [1984], the geometric school of thought about brain function, including coordination, reaches back about 400 years to Descartes [1629]. Descartes' insert in Fig. 1 [ab12], from Pellionisz [1984], illustrates his most reasonable idea – in retrospect – that by the Cartesian coordinates both key concepts of living systems were comprised. Both multimodal composition of entities was shown, as well as a functional recursion of information; see the finger-movement under the feedback of visual control. TNT "simply" generalized the Cartesian x, y, z (and t) coordinates of the Minkowski-spacetime manifold, where it became evident that in non-orthogonal expressions generalized vectors (tensors) profoundly differ if expressed in a "sensory or motor manner".

The tensorial scheme in Fig. 1 uses a (minimal) 2-component sensory vector and a higher (3) dimensional motor vector that expresses the same physical object (invariant, in his case a displacement).The scheme illustrates the contravariant motor efferent vector, as well as the recursing covariant proprioception vector. As explained in detail in Pellionisz [1984] (see also Fig. 2 here) this recursion converges in the brain stem in the Eigenvectors that are essential to build the cerebellar metric, as the matrix-product of Eigenvectors, found by recursive oscillatory tremor. Even this schematic representation points out that the interim oversimplification (e.g. exemplified

by Lorente de No [1933] that 3-neuron reflex arcs carry a one-to-one representation) is mistaken; Szentágothai [1949]. Single “loops”, e.g. reflex-arcs are surpassed by a many-to-many network of neural nets, harboring some massive interconnections, described by vector-matrix tensor geometry.

Switching to the seminal work of Genomics [Mendel 1866] regarding many phenotypes (he investigated 7 characteristic inheritable traits, in parallel), a similar one-to-one oversimplification ensued; a decided effort to associate with, or rather, to pin “one phenotype on a single genotype”. A key message of this chapter is to draw a parallel that the “single gene-to-single phenotype” approach is likewise futile; as is a “loop-type” single reflex. Instead, there is a “neural network”-type “many-to-many” interaction among, say n, phenotypes, and the underlying, say k, genotypes. It is strongly believed that the cerebellum, with its already modeled multi-component factors is the best platform to sort out the underlying mathematics of *“multicomponent dual representation of covariant and contravariant functors (defined as objects that relate categories)” of not only the cerebellar neural networks, but also of their genomic roots.* In order for this to happen, science needs to specify the mathematics that underlie “a biological system theory”; Bertalanffy [1934]. Identification is essential for both the neural network and for the underlying genome, including the suggestion here that they are conceptually identical.

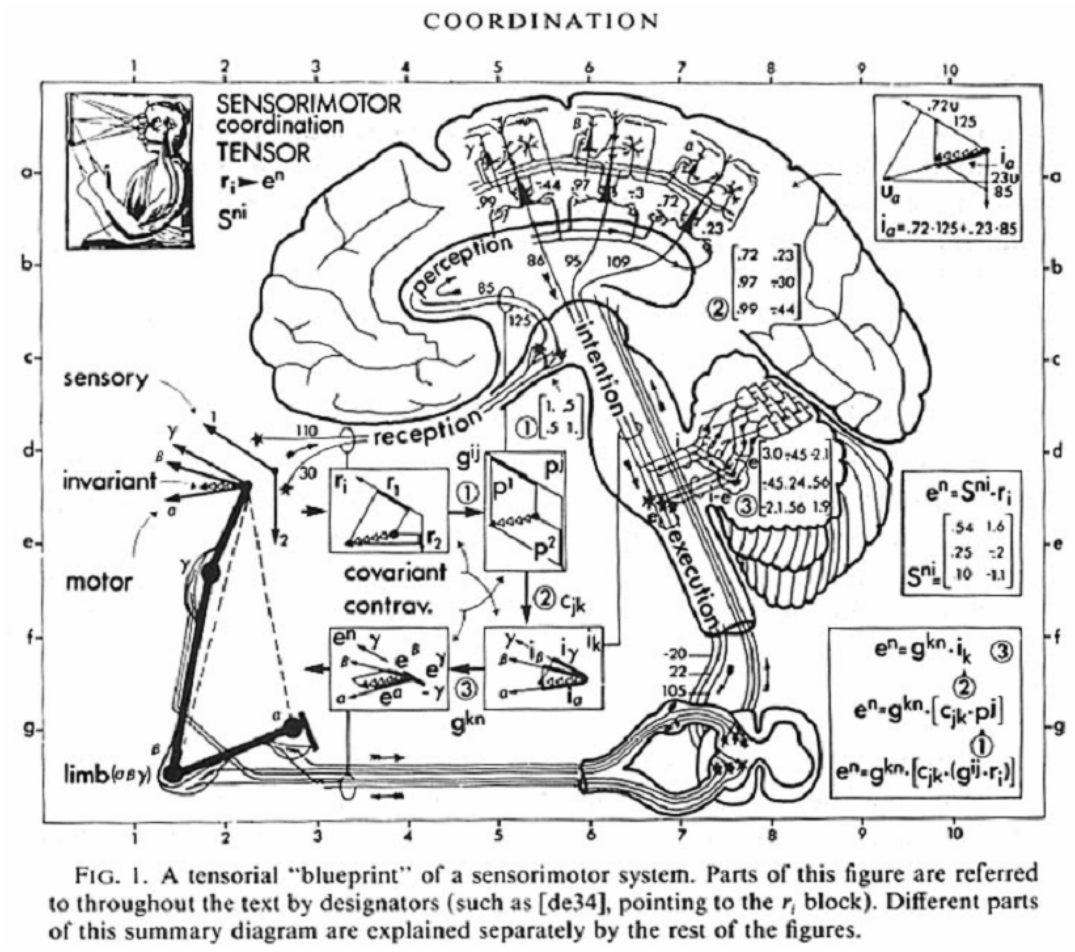

FIG. 1. A tensorial “blueprint” of a sensorimotor system. Parts of this figure are referred to throughout the text by designators (such as [de34], pointing to the r_i block). Different parts of this summary diagram are explained separately by the rest of the figures.

Fig. 1. An example of specific “System Theory” identifying the modern mathematics of Descartes’ classic concept of “coordinates” [ab12]. *Surpassing the Cartesian frame of rerence by generalized coordinate systems used by Nature for sensorimotor coordination [de13], cerebellar coordination is explained in terms of tensor geometry (Fig. 1 from Pellionisz [1984]). For biological organelles, organisms and organs, in this case that of cerebellar sensorimotor system, no “Biological System Theory” will be “software enabling” unless the intrinsic mathematics is identified, as it is shown here, or better. Further explanation is in the text, and the mathematical procedure is elaborated in Pellionisz [1984].*

Further, as it is suggested here, science needs to move away from a “one-to-one” and “arrow-type” mapping towards the “many-to-many more” and “recursive” *and* dual representations. This is important not just for theory, but for entire industries. The “Big Pharma” model of *“one gene, one disease, and one billion dollar pill”* is obsolete for over a decade because of a simplified and incorrect “one-to-one” assumption. Now the future lies in the generalization of covariant- and contravariant neural network representation for the genome-epigenome (hologenome) system. The cerebellar biological neural networks, as shown, provide a precedent for this mathematical insight that is also applicable to genomics.

3. Generalization of Recursion from Cerebellar Neuroscience to Genomics; Covariant and Contravariant RNA Functors and their Eigenstates

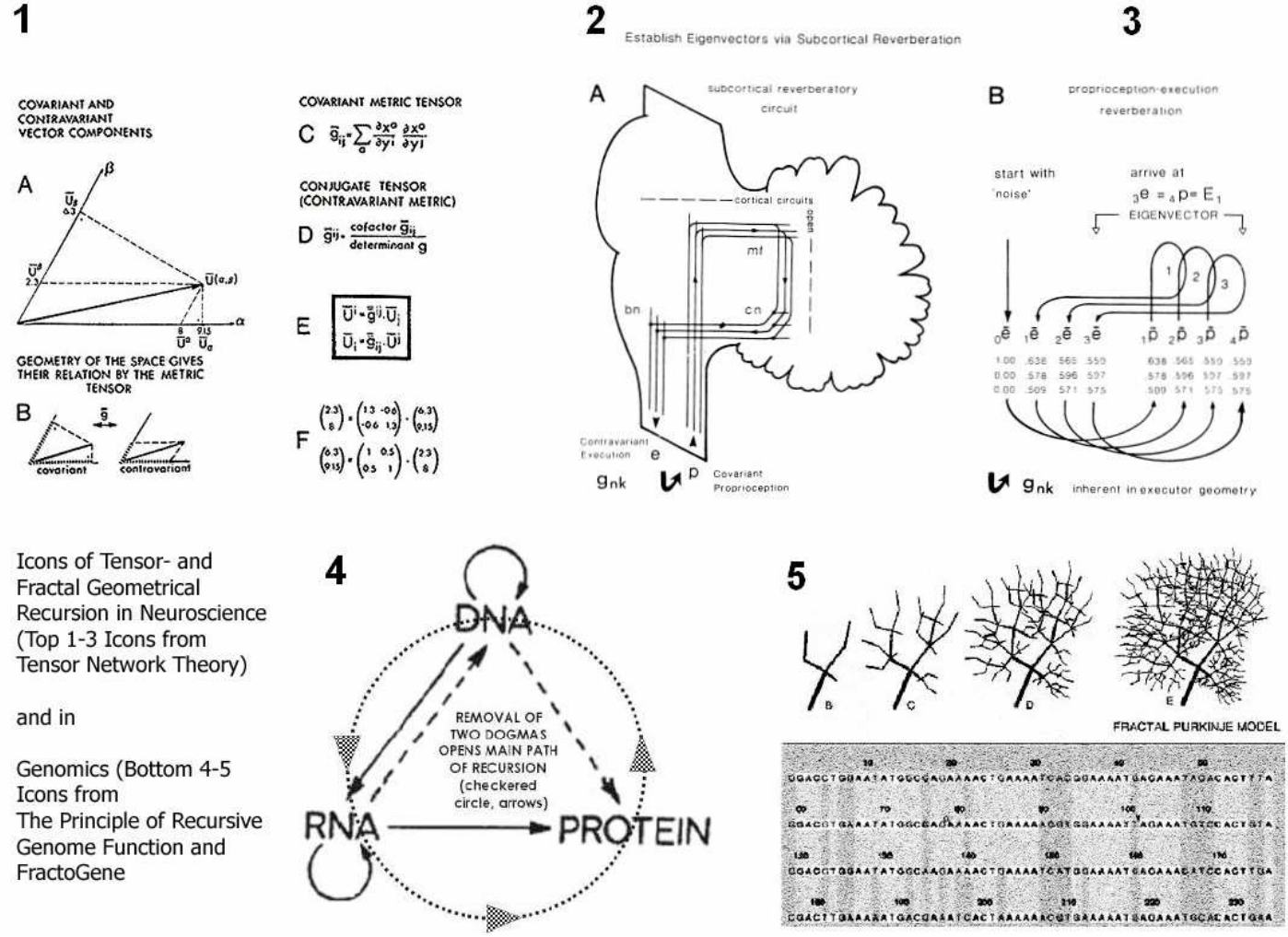

Fig. 2. Icons of Tensor- and Fractal Geometric Recursion in Neuroscience for a Geometric Unification of Neuroscience and Genomics using cerebellar cellular and network systems. *For originals and explanation in detail see neuroscience Icons (from TNT) for Icon 1 Pellionisz and Llinás [1980], 2-3 Pellionisz and Llinás [1985]. For the bottom (genomics) Icons 4-5, for Icon 4, see Pellionisz [2008a,b], and for Icon 5 see Pellionisz [2002, 2003] and Simons and Pellionisz [2006a]. Icon 1. comprises the dual valence of vectors, if using non-Cartesian (generalized, non-orthogonal) coordinate systems. The covariant and contravariant vector components are shown in Panel A together, while in Panel B separately. It is cardinal that they are not the same either in their values or in the way how they represent the invariant, the covariant components can be independently measured, but they do not add physically, while contravariants are interdependent, but they do physically generate the invariant. Panels C-F in Icon 1 show that the covariant and contravariant vectors can be converted to one-another by the metric tensor. Icon 2. Panel A shows that the covariant proprioception vectors could recur through the brain stem (even without a cerebellum or sensorimotor cortex) via mossy fibers to cerebellar nuclei, and could be directly (but inappropriately) used as if they were true contravariant executor vectors. Thus, a recursion may take place, as shown in Icon 3. Close inspection shows that in a non-orthogonal system of coordinates, starting with any vector (even a noise vector), after several recursion the co- and contravariant expressions converge into the Eigenvectors (where the incoming and outgoing vector components are the same). Physically, this mechanism is an uncontrolled but convergent tremor, mathematically the discovery of Eigenvectors. Dyadic products of Eigenvectors, yielding a matrix, create the metric tensor. Icon 4 shows The Principle of the Recursive Genome Function, that permits a DNA>RNA>PROTEIN>DNA… recursion (after discarding the obsolete notions of Central Dogma and Junk DNA). Note, that for the purposes of simplicity Icon 4 shows the recursion as a single circular line; however it symbolizes multi-component (vector) entities. The cardinally important Generalization of Recursion (from neuroscience to genetics) is the concept introduced here that the coding DNA vectors (many "exons" acting together), when transcribed, create RNA vectors that are of contravariant valence, since their translation into protein vectors creates physical objects. However, when protein vectors are signaled (measured) by non-coding DNA via bonding not only to homeodomains but to ncDNA vectors, are transcribed into another RNA vector, this time of covariant (sensory) valence. Thus, a recursion, similar to one shown in Icon 3 converges into the Eigenstates of the recursion in the genome, and the cRNA and ncRNA Eigenvectors produce the metric, comprising the functional geometry of the genome function. If the recursion converges to follow the Weyl' Law on Fractal Quantum Eigenstates, the genomic recursion switches the growth of fractal protein structures (such as a Purkinje neuron, shown in Icon 5) into the next step of recursive hierarchy. The physiological process requires canceling (methylating) ncDNA segments perused in the recursion, see Fig. 9, such that the ncDNA fractal segments, governing growth according to FractoGene are not overused. It follows, that hypo-methylation and incorrect chromatin modulation could permit an uncontrolled (cancerous) growth as shown in Fig. 9 (yellow "cookie"). For further details, consult the original papers containing Icons 1-5 and the text of this chapter, relating the seminal concept of generalization of recursion described in Fig. 2 with the fractal recursive iteration shown in Fig. 9.*

As it was shown over three decades ago in Pellionisz and Llinás [1980] if using non-Cartesian (generalized, non-orthogonal) coordinate systems, see Icon 1 in Fig. 2, invariants (such as displacement) are represented in with a dual valence. The orthogonal projection-components, named covariant tensor-components in mathematics by Sylvester [1853], can be independently established, however, covariant components do not physically assemble the invariant. In turn “motor expressions”, expressed as interdependent parallelogram-type coordinates, that he called contravariant tensors, do assemble the object in a physical manner. It is cardinal in mathematics of generalized coordinates (tensor geometry) that a matrix can convert the “covariant sensory intention vectors” into “contravariant motor execution vectors”; see Panels C-F of Icon 1. The matrix that does this is the many-to-many interconnection-system of a massively parallel neural network of the cerebellum. Thus, the cerebellar sensory-motor coordination is accomplished by the conversion via the metric tensor. The metric comprises the geometry of the non-Cartesian multidimensional space-time, embedding both sensory and motor events. This perhaps difficult but cardinal concept of sensory- and motor components as co- and contravariant vectors was most lucidly encapsulated by Anderson [1990] pp. 351-355, in Anderson et al, [1990].

Dual, covariant and contravariant functors, shown in Fig. 2 Panel 3, if they are freely let to recur (when proprioception vectors are directly used by recursion as if they were execution vectors, without the cerebellar cortex, see Icons 1-2 in Fig. 2), converge into the Eigenstates (where the normalized covariant- and contravariant representations are identical – while in general they are different). Finding the Eigenvectors characterizing the Eigenstates by free recursion (that in sensorimotor systems is manifested in uncoordinated, oscillatory movements) is essential, since the metric tensor (and its inverse, or Moore-Penrose Pseudoiverse for overcomplete space Pellionisz [1984], capable of converting covariants to contravariants, is obtained as a matrix-product of the Eigenvectors).

In category theory covariant- and contravariant- as well as mixed valence of functors (vectors and generalized vectors, tensors, are just one specific type of functors; they relate invariants to coordinate axes) are both well established and reaffirmed; Francis [2008]. Herein, with the conceptual guidance of Icons 3-4 in Fig. 2, the dual representation is generalized to the interpretation of RNA system in Genomics. There is no debate that so-called amino-acid-coding RNA-s (cRNA-s, as a multicomponent, vectorial entity) physically aggregate physical objects (proteins). Thus, the valence of cRNA-s is contravariant, similar to motor vector components that also have to assemble the physical object. The contravariant cRNA vectors, however, via RNA self-replication; Glasner et al. [2000] are available not only to construct proteins, but to interact (interfere) with the rather different covariant multicomponent ncRNA vectors. These measurements, what protein-systems are already built, arise when proteins bind to non-coding DNA (both in intergenic and intronic sequences) involving transcription factors; Kornberg and Baker [1992]. Through the arising ncRNA functors (multicomponent vectors), the already built proteins are thus “measured” not just by a single sequence, referred to as “homeoproteins” generated by a “homeodomain”; Foucher et al. [2003], but a single protein-component is signaled by the many components of even a single but multi-component ncRNA covariant vector. Compare the concept to covariant sensory vectors providing independent measures of motor events in Icon 1 of Fig. 3.

Putting the RNA system here into a new conceptual framework, also re-defining the role of intronic and intergenic “non-coding” (formerly, “Junk”) DNA, recalls earlier metaphors. Interpreting the RNA system as a “hidden layer”, an implication referring to interconnections known in neural nets for decades; Mattick [2005] phased out his earlier metaphor that conceptually compared the RNA system to the man-made “operating system of computers”; Mattick [2001]. Recently, even “genomic matrix” relating to fractals and chaos; Petoukhov and He [2010] and even “RNA matrix” approaches emerged; Izzo et al. [2011]. However, the co- and contravariant valences of RNA functors have not been recognized to date. This generalization of valence of functors from sensory- and motor vectors to covariant as well as contravariant RNA multicomponent entities provides an opportunity to approach the role of RNA systems in coordinated genome function in a novel manner; that is both conceptually and mathematically already identified in living systems (cerebellar neuronal networks) that the genome-epigenome system is known to generate.

Appreciation of the valences of RNA functors opens new vistas beyond approaching the RNA-metric from a mindset that moves the perspective of science beyond man-made technologies like operating systems of computers [Mattick 2001]. Looking at the RNA system in a new light as “the metric tensor of protein building genic sequences regulated by protein sensing non-coding sequences”, the RNA system is conceptually likened to a “genomic cerebellum”. First of all, this permits deploying already proven advanced geometric (thus software-enabling) analysis of experimental results of genomics. Secondly, the perspective on evolution is affected by recalling the shifting metaphor by Mattick from “operating system” to “hidden layer” [2001 versus 2005] and his reference that the RNA system serves a “coordinated genome expression”. One cannot help noticing that the “invention” through evolution of the physically separate, additional cerebellar neural network (with the shark) provided for a new class of more highly coordinated vertebrates. The conceptual equivalence is noted, therefore, that much of single-cell organisms contain a minimal amount of “non-coding DNA” – thus appear to operate with minimal covariant ncRNA; similar to organisms before the cerebellum appeared, permitting only an imprecise, un-coordinated execution of genomic commands. As the amount of non-coding (regulatory) sequences hyperescalated, the emerging RNA-metric permitted the coordinated growth and governance of complex (also multicellular) organisms. This new interpretation of the RNA system is to be compared to Mattick’ referral to “the Cambrian explosion”; Mattick [2004].

At the least, identification of a common advanced geometry intrinsic to living systems makes “System Theory” approaches to genomic systems mathematically explicit. A more remote but an inevitable goal for the use of a common advanced geometry is to accelerate the unification of genomics and neuroscience. It is fully realized that building this seminal idea into a robust school of thought will require significant time and resources.

3.1. Recursive Algorithms Rule both Vector-Matrix and Fractal Representations

Algorithms based on recursion, see Icon 4 of Fig. 2 from Pellionisz [2008a, b] share the fundamental property that each state of the system is deduced from its previous states. Recursion, in itself, does not discriminate analogue (e.g. traditional feedback) mechanisms from digital deduction as, e.g. in the sequence of Fibonacci-numbers, where each subsequent integer is the sum of the previous two. The metric tensor characterizes the non-Euclidean geometry with integer dimensions, established by recursion of covariants to contravariants to compose the metric from Eigendyads, Pellionisz and Llinás [1985]. The embedding Minkowski spacetime manifold, however is “smooth”, mathematically speaking it is derivable. However, Purkinje neurons show a non-Euclidean, moreover, a discrete geometry with fractal (non-integer) dimension; Pellionisz [1989].

Realization that the same cerebellum utilizes recursion of dual vectors, as well as its main type of neurons, the Purkinje cells are built by an also recursive, but by a rather different fractal iterative recursion; see Icon 5 of Fig. 2; Pellionisz [2002, 2003] a cardinal question arose ever since the fractal model of Purkinje neurons; Pellionisz [1989]. The question became even more vexing with the FractoGene concept; stating that fractal DNA governs growth of fractal organelles such as the Purkinje neuron, fractal organs such as the lung, circulatory systems and organisms such as the Cauliflower Romanesca pictured in Pellionisz [2008]. The question was if the vector-matrix and fractal representations are in a mathematical conflict with one-another, or rather, if they reveal another profound dualism, similar to one already encountered in physics.

The question was also conceptual regarding not only the mathematics, but also possibly referring to a “language”. The “early wave” of looking at fractality of DNA suspected it as a “language”; Flam [1994]. The concept of a “language”, however, does not appear to be consistent with the concept of “sensorimotor coordination”. Resolution of the question became easier once the “hint” that fractality reflects a “language” was dismissed; Chatzidimitriou-Dreismann [1996]. Section on the Zipf-Mandelbrot Parabolic Fractal Distribution

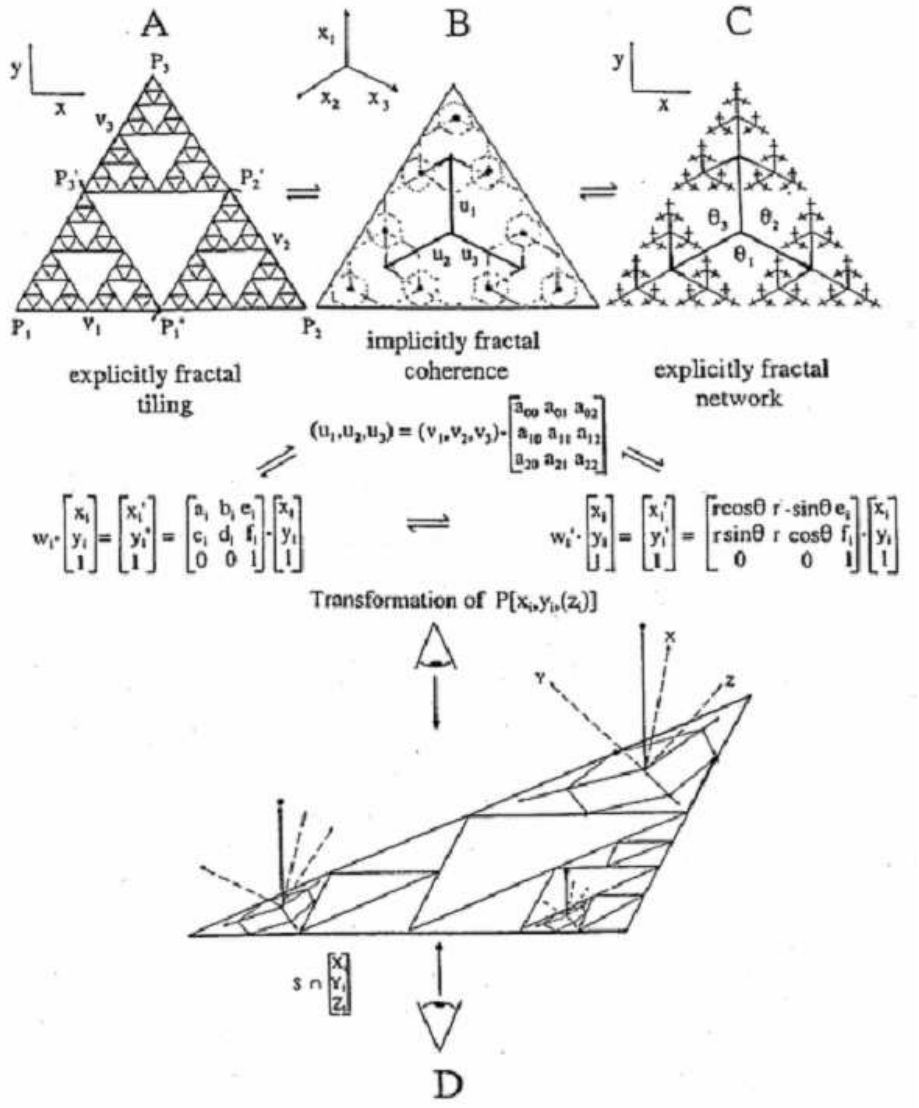

Fig. 3. The conceptual linchpin of Multidimensional Generalized Vector-Matrix approach (D) to the Fractal Approach (A,B,C), *from Bieberich [1999].*

shows below that the established fractality of the genome conceptually supports FractoGene; fractal growth of Purkinje cells governed by fractal DNA. Both in the DNA and in networks of neurons the fractality characterizes the geometry in a consistent manner.

The question was settled by Bieberich [1999], see his Figure reproduced as Fig. 3 of this chapter, to show a conceptual consistency of fractal and vector-matrix representations. For a general audience, the connections between fractals and multiplication of matrices was also shown by Gazalé [1999]. Thus, a geometric characterization of sensorimotor function and the geometry of the Purkinje neurons that implement smooth (derivable) function by non-derivable fractals are not only compatible, but mutually convertible. The revelation by Bieberich [1999] was not entirely surprising; given the known fact in physics that light can be seen as a wave-phenomenon, or particle-phenomenon, depending on the theory of Schrödinger or Heisenberg. Thus, the Biebererich-diagram is intellectually rather pleasing. Even more intriguing is its extension towards fractal internal representation (consciousness) in Bieberich [2011].

Based on insights from fractal modeling of Purkinje neuron; Pellionisz [1989], utilities could be developed based on the of fractality of both DNA and the organelles, organs and organisms grown by the genome; FractoGene by Pellionisz [2002, 2003, 2006a,b]. The FractoGene algorithmic approach to the whole genome provided quantitative predictions that could be verified or refuted by experimentation, moreover the *"Fugu Prediction of FractoGene"* (that the 1/8 of the noncoding DNA of fugu compared to that of the human should result in a "fractal primitive" dendritic tree in the fugu), was supported by experimental results; Simons and Pellionisz, [2006a,b].

3.2. Tensor Network Theory: Vector-Matrix Recursion as Basis of the Cerebellum Acting as a Sensorimotor Metric Tensor

Recursion of sensory to motor vectors (and the generalization of variance of RNA functors) was characterized by Icon 2 of Fig. 2 as an essential procedure to converge into Eigenvectors, with their matrix-product comprising the geometry in the metric tensor. With the example of encyclopedic Fig. 4 of this chapter from Pellionisz [1987] it is shown how such metric is the basis of an entire system of gaze control; stabilizing the head by the vestibulocollic sensorimotor neural network.

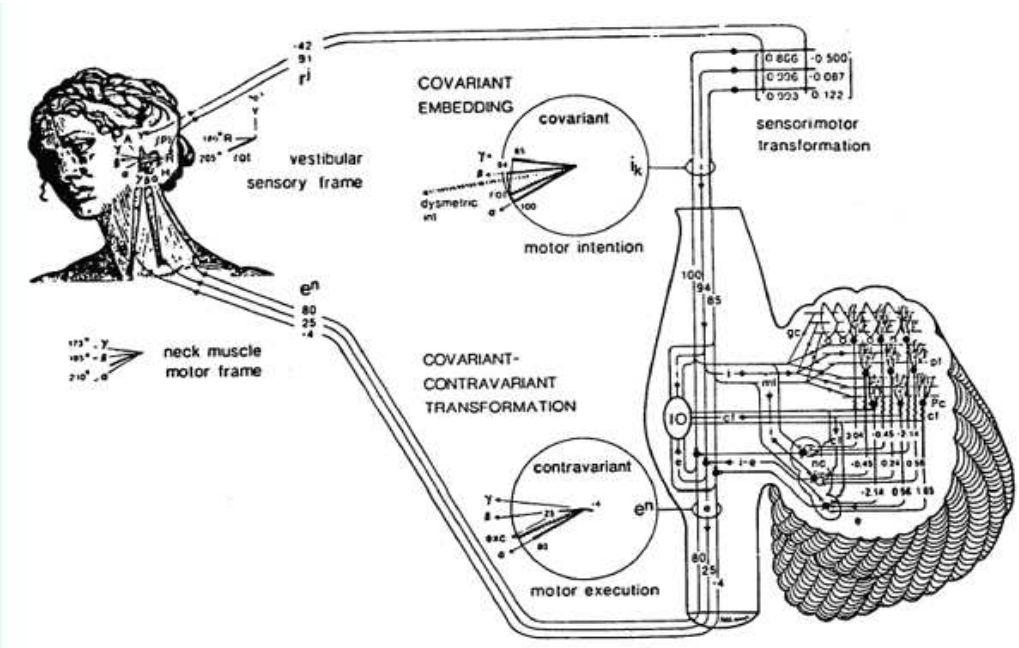

Fig. 4. Tensor network model of the vestibulocollic reflex, embodying a covariant intention to contravariant motor execution transformation via the cerebellar neuronal network. *From Pellionisz [1987]. For details, see the original publication and the text below. This Figure also serves as the inspiration of the seminal concept of generalization of TNT to Genomics. The parallel lies in the fact that a physical object of the head movement is both measured by the covariant sensory vector that converted both in dimensionality and co- to contravariant valence. Likewise, the genome expresses physical objects (proteins) both by protein-coding codons (in a contravariant manner), that can be measured by similar (but non-coding triplets, wherein the detection is covariant), but in order to attain a quantum fractal eigenstates of stable protein systems a many-to-many RNA converter is needed. This is conceptually similarly to the sensory-motor transformer of the cerebellum.*

Icons 2-3 of Fig. 2 showed that sensory functors could recur directly, used in an unchanged manner, as motor functors. However, the recursion would result in an oscillation converging into Eigenvectors. In the cerebellar sensorimotor system, the Eigenvectors are

imprinted in the inferior olive; Pellionisz and Llinás [1985]. In turn, as shown in Fig. 4 here, Eigenvectors from the inferior olive give rise to their matrix-product implemented by the neuronal network of cerebellar cortex. The scheme shown in Fig. 4 stabilizes gaze (head position) by a two-step operation. First, there is a covariant embedding from a symbolically 2-dimensional sensory vector into an also covariant, but higher (figuratively, 3) dimensional motor intention vector (i) – that would go directly to (mis)serve as an imprecise execution vector (since motor vectors must be contravariant; (i) should be (e). Through the ascending mossy fibers, the (i) covariant intention vector is both converted into the (–e) contravariant vector (negative, since Purkinje cells are inhibitory), that with the mossy fiber collateral (i) vector in the cerebellar nuclei constitutes an output vector (i-e). Thus, the brain stem would send out instead of the covariant intention vector (i) the proper e=i-(i-e) precise contravariant execution vector. This architecture explains why the entire sensorimotor would work (as for a dysmetric patient; even Purkinje cell affected only by alcohol) with intentions directly executed, but the additional neural network that was a nifty improvement as an addition to the brain of the shark, makes a dysmetric direct execution of intentions into one that matches the physical geometry of the executor system (in this case, muscles) with its internal metrical representation.

Coordination of an entire sensorimotor architecture is presented here to illuminate how non-trivial the generalization of contra- and covariant cRNA functors directly recursing into ncRNA covariant functors is. Such direct recursion is excellent for finding the Eigenstates of a DNA>RNA>PROTEIN recursive system, but the multicomponent RNA Eigenvectors must interact in an all-to-all manner, by means of RNA interference, not just of one component, Fire et al [1998], but in a many-to-many multicomponent manner. Also, the sensorimotor coordination scheme is to illuminate why RNA interference is “silencing” – conceptually similarly to the inhibitory effect of cerebellar Purkinje cells.

Development of the school of functional geometry of a comprehensive system of coordinated genome function, comparably to that of a sensorimotor apparatus, requires a long-term program. One of the most difficult questions is if the genomic recursion obeys the Fractal Weyl’s Law on Fractal Quantum Eigenstates, see Shepelyansky [2008], originally Weyl [1912]. This question will be discussed in the section “Future Directions”.

3.3. Fractals are Pervasive in Nature; both the Cerebellar Brain Cells and the DNA are Fractal Objects

Mandelbrot [1983] coined the term “fractal” in his epoch-making book only about a quarter of a Century ago, but the impact of identifying fractal geometry intrinsic to Nature is already profound.

3.3.1. The Zipf-School Suspected that the DNA contained a Fractal Language

The first “hints” that the A, C, T and G nucleotide-sequences of DNA (especially of non-coding DNA) possibly harbored a (mathematical) “language” was published before the epoch of “massive whole genome sequencing”, in 1994 in Science, see Fig. 1 in Flam, [1994]. Its original captation: *“Line of evidence. Plotting frequency against rank of arbitrary ‘words’ in noncoding yeast DNA yields the linear plot found in human language”* reveals the key word “arbitrary”. Note that “words” of the non-coding DNA were 3-8 bases, sampled in an unjustified manner. Neither graph appeared to conform to the straight “Line of evidence” of Zipf’s law.

The study reported by Flam was based on a comparison with the empirical “Zipf’s law”, that applies to natural languages Zipf [1949]. The distribution of frequencies (actual occurrences) of words in a large corpus of data versus their rank is generally a power-law distribution, with exponent close to one. Zipf's law is thus an experimental law, not a theoretical one. Zipf-like distributions are commonly observed, in many kinds of phenomena. However, the causes of Zipf-like distributions in real life are a matter of some controversy, with DNA being no exception.

While the early observations applied to DNA in 1994 were found worthy of reporting in Science and were widely heralded that “something interesting was lurking in the junk [DNA]” the “Zipf-test” was inconclusive. Review by Simons and Pellionisz [2006a] pointed out that investigators failed to detect "well-defined scaling or fractal exponents"; Chatzidimitriou et al, [1996] or “any signs of hidden language in non-coding DNA"; Bonhoeffer et al, [1997].

Empirical law aside, the biggest problem was the definition of "words" in the DNA. First, Harvard linguistics professor Zipf (1902-1950) established his “law”, based on observations on the English language, in which “words” are taken for granted. He found that in text samples the frequency of any word was roughly inversely proportional when plotted against the rank of how common each word was; the frequency of the *k*-th most common word in a text was roughly proportional to 1/*k*. Plotting both frequency and rank on a logarithmic scale, “Zipf’s law” was expected to yield a declining linear graph also for “words” of the DNA.

When applying this natural language lingustics to DNA the results were not entirely convincing (Fig. 1 of Flam, 1994) . The problem was not only that the graphs did not quite conform to the linear Zipf’s law. It is unacceptable that the definition in the noncoding DNA was completely and explicitly *arbitrary*. Of course, there was no definition at that time of what A, C, T andG strings might constitute “words”. In the analysis conducted by Mantegna et al [1994]: *“when the group arbitrarily divided up their samples of junk [DNA] into “words” between 3 and 8 bases long and applied the Zipf test, the telltale linear plot emerged”*.

Looking at the reproduced Fig. 1 of Flam (1994), the plot (for non-coding DNA “words” open squares on a log-log scale) starts fairly close to linear, but drops off remarkably at the tail end. The original Flam-diagram of the Zipf-law for DNA was even more controversial when it was applied to the “coding regions” of the DNA (see graph of open circles in Fig. 1 from Flam, 1994). Here, Flam claimed that

the Zipf-law "failed" – and the reason cited was that *"The coding part [of the DNA] has no grammar – each triplet of bases corresponds to an amino acid in a protein. There's no higher structure to it".*

Today, both the "definition" of arbitrarily picked 3-8 letter strings for "words" and the "axiom" that there is no higher structure to coding DNA appear demonstrably dogmatic.

Zipf's law is most easily observed by scatterplot the data, with the axes being log(rank order) and log(frequency). The simplest case of Zipf's law is a "1/*f* function". Given a set of Zipf-like distributed frequencies, sorted from most common to least common, the second most common frequency will occur 1/2 as often as the first. The n^{th} most common frequency will occur 1/*n* as often as the first. However, this cannot hold precisely true, because items must occur an integer number of times: there cannot be 2.5 occurrences of a word. Nevertheless, over fairly wide ranges, and to a fairly good approximation, many natural phenomena obey Zipf's Law.

3.3.2. The Genome is Fractal: Grosberg-School Suspected that the DNA Showed Fractal Folding

The classic book of the mathematician who coined the word "fractal" (as a measure of dimension of roughness of results of recursive procedures), Mandelbrot [1983] generated a huge impetus into the direction of pulling away from looking at the genome as a language, and looking at fractals more as the "geometry of nature". The twin schools of thought, towards approaching the structure of the genome – and the protein-structures whose development it governs, manifested in the seminal work by Grosberg et al [1988, 1993] to claim that the folding of DNA strands were fractal. Decades later, as an eminent example how established methods of biochemistry can be used to support paradigm-shifts, the Science cover article appeared [Erez-Lieberman et al. 2009], in effect the Science Adviser to the US President, Eric Lander appealing *"Mr. President, the Genome is Fractal!"* Inspired by the Hilbert-curve, a recursive folding that provides the much needed propensities. First, it is knot-free to permit uninterrupted transcription. Second, it is ultra-dense to enable squeezing the 2m-long DNA strand into the nucleus of a cell with 6 micron diameter. Remarkably, the Hilbert-curve is capable of filling the entire space available, in its 3D form its fractal dimension is 3.0. Third, it also provides the advantage that is paramount for The Principle of Recursive Genome Function, Pellionisz [2008a,b] that the DNA can be read not only serially, from one end to the thread to the other, but because all segments of the DNA are in maximal proximity to one-other, they can also be read in parallel.

3.3.3. The Perez-school shows that the DNA is Fractal at DNA, Codon- and Full Chromosome Set and whole Genome Levels

The Perez-school of study of recursive systems was interdisciplinary [Perez 2011b] and showed first results in 1988 [Perez 1988a and 1991].The fractal nature of A, T, C and G coding or non-coding nucleotide sequences, chromosomes and genomes was evidenced over two decades, see review Perez [2011a]. Details, e.g. Perez [1991] and Marcer [1992] are comprised in two books; Perez [1997 and [2009a]. The results spanning from recursive studies through DNA and full genome analysis, including full set of chromosome levels, Perez [2008] are likely to be a serious candidate to the measure of "Abstract DNA Roughness" as proposed in Section 5.2.

3.3.3.1. Fractals to DNA numerical decoding: towards the Golden ratio. "*Small is beautiful".* Inspired by the recursive "Game of Life"; Gardner [1970] using the largest computers in the time a cellular automata a large random 0/1 cell populations was run in 1988 [Perez 1988a and 20089b. After 110 parallel network iterations, with a recursive single-line code, a "clown" pattern (see Panel 1 of Fig. 5) emerged from the small 7 cells "U" (see upper left corner of Panel 1 of Fig. 5 from Perez [1988a]). A strong illustration of « small is beautiful » is the discovery of a predictive formula of the Mendeleev's Elements periodic table architecture, Perez [2009a and 2009c].

3.3.3.2. The "Fractal Chaos" artificial Neural Network. In the eighties, various parallel artificial neural networks were explored Perez [1988a 1988b], with a particular interest in discrete waves and by fractals. The fractal chaos is summarized by right-bottom Panel 5 of Fig. 5. In the dynamics of the fractal, a curious focal point seems to emerge: the "Golden ratio". The fractal network also provides "déja vu" recall memory and holographic-like memory, Perez [1990a and 1990c]. At that time chaos in the DNA was also searched, but it is discrete; A, T, C and G bases could be coded by integers while chaos theory is based on real numbers. Note that the ratio between 2 Fibonacci integers is near to the Golden ratio. This raised the question of an integer-based chaos theory. Indeed, a hyper-sensitivity of the fractal for inputs based on recursive Fibonacci numbers was demonstrated; Perez [1990b].

3.3.3.3. "DNA SUPRACODE" overview. A connection between DNA coding regions sequences as gene sequences A, T, C and G patterned proportions and Golden ratio based Fibonacci/Lucas integer numbers was proposed; Perez [1991], Marcer |1992], see also Fig. 5. Panel 2. Correlation samples were established in genes or gene-rich small genomes with evolution or pathogenicity (example of HIV genome particularly; see the book Perez [1997]). "Resonances" were analyzed, where a resonance is a Fibonacci number of contiguous A, T, C and G nucleotides (i.e.144). If this sub-sequence contains exactly 55 bases T and 89 bases C, A, or G, this set was called a "resonance". Thousands of resonances were discovered (see upper right corner of Panel 2 of Fig.5. from Perez [1991]) : in HIV -the whole genome is long of about 9000 bases-, there are resonances overlapping about 2/3 of the genome.

3.3.3.4. In single-stranded DNA Human genome, codons population are fine tuned in Golden ratio proportions. A new Bioinformatics bridge between Genomics and Mathematics emerged; Perez [2010]. This "Universal "Fractal Genome Code Law" states that the frequency of each of the 64 codons across the entire human genome is controlled by the codon's position in the Universal Genetic Code table. The frequency of distribution of the 64 codons (codon usage) within single-stranded DNA sequences was analyzed. Concatenating 24 Human chromosomes, it was demonstrated that the entire human genome employs the well-known universal genetic code table as a macro structural model.

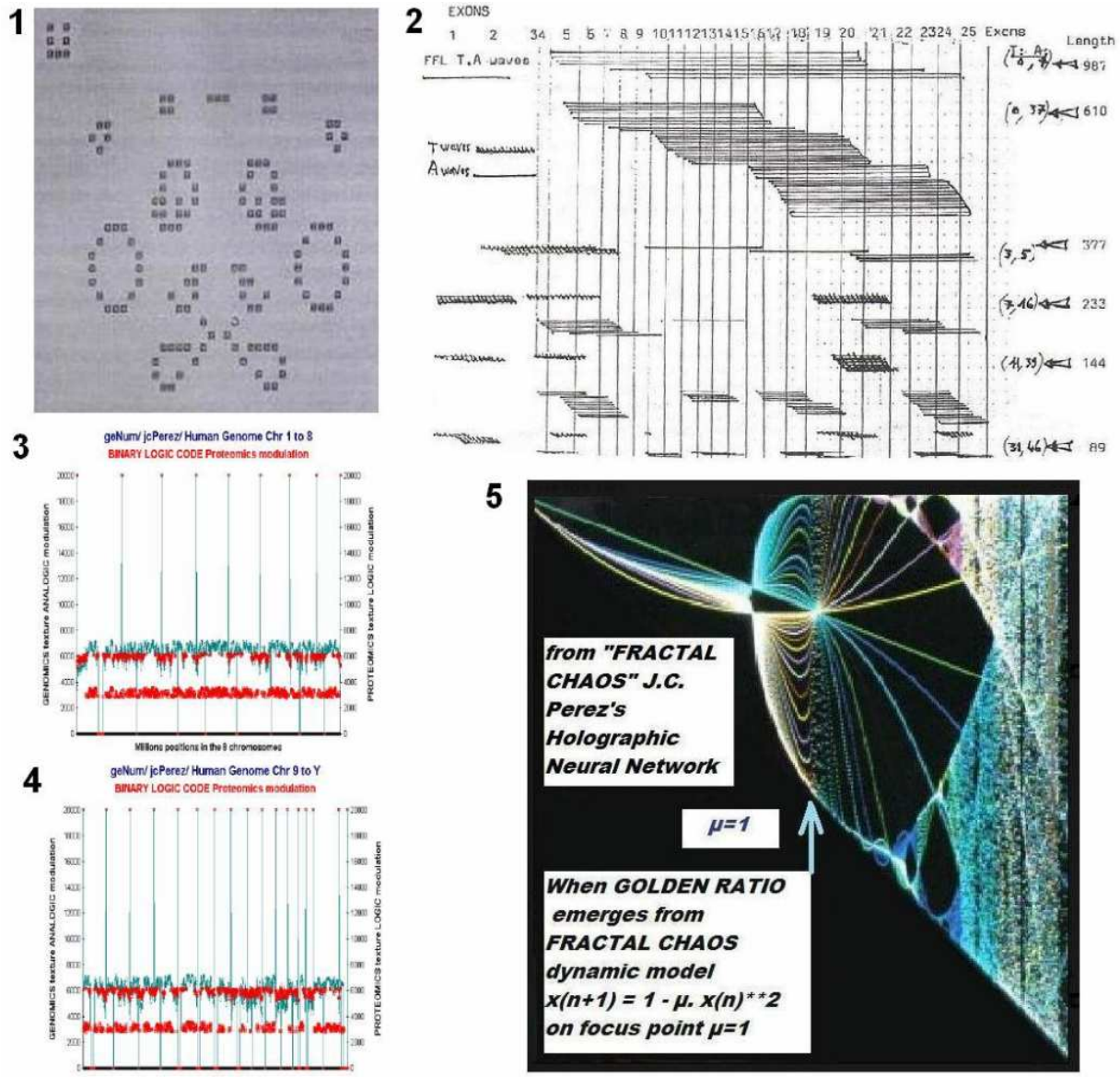

Fig. 5. Examples from the Perez School of Recursive Results. *Panel 1: "Clown" emerging from U (upper left corner), citing original recursions in 1988 by Perez. Reproduced from Perez [2009b]. Panel 2: DNA supracode and recursive Fibonacci series: 1 1 2 3 5 8 13 21 34 55 89... Example of resonances in HUMC1A1 gene. Reproduced from Perez [2011a]. Panel 3: Chromosome 1-8. The Evidence of Binary Proteomics Code (red) and Modulated Genomics Code (blue) at the Whole Human Genome Scale. Green: Genomic, Red: Proteomic. Reproduced from Perez [2011a]. Panel 3: Chromosome 9-Y. Reproduced from Perez [2011a]. Panel 5: Perez [2010] Fig. 5. Fractals to DNA numerical decoding: the Golden ratio. Evidence of Golden ratio hypersensitivity in a specific region of the "Fractal Chaos" recursive.*

The position of each codon within this table precisely dictates its population. So the Universal Genetic Code Table not only maps codons to amino acids, but serves as a global checksum matrix. Frequencies of the 64 codons in the whole human genome scale are a self-similar fractal expansion of the universal genetic code. The original genetic code kernel governs not only the micro scale but the macro scale as well. Particularly, the 6 folding steps of codon populations modeled by the binary divisions of the "Dragon fractal paper folding curve" show evidence of 2 attractors. The numerical relationship between the attractors is derived from the Golden ratio. It was demonstrated that:

(i) The whole Human Genome Structure uses the Universal Genetic Code Table as a tuning model. It predetermines global codons proportions and populations. The Universal Genetic Code Table governs both micro and macro behavior of the genome.

(ii) The Chargaff's second rule from the domain of single A, T, C and G nucleotides was extended to the larger domain of codon triplets.

(iii) Codon frequencies in the human genome were found to be clustered around 2 fractal-like attractors, strongly linked to the Golden ratio 1.618; Perez [2010].

3.3.3.5. A strange Meta-Architecture organizes our 24 Human Chromosomes. A curious interaction network was found among our 24 human chromosomes Perez [2011a], see Fig. 5, Panels 3-4 for human Chromosomes 1-8 and 9-Y, respectively. It was proven that the entire human genome codon population is fine-tuned around the *"Golden ratio";* Perez [2010]. Across the entire human genome, there appears to be an overall balance in the whole single-stranded DNA. This digital balance fits neatly around two attractors with predominant values of 1 and (3-Phi)/2, where Phi is the Golden ratio. Yet, the same analysis applied individually to each of the 24 chromosomes of humans and to each of the 25 chromosomes of the chimpanzee which reveals a 99.99% correlation between both genomes but diversity and heterogeneity particularly in the case of our chromosomes 16 17 19 20 and 22; see the book *"Codex Biogenesis"*, Perez [2009a]. Thus, a paradox emerges. The same analysis shows a global unity across the genome, whereas, applied to each of the constituent chromosomes of this same genome a great heterogeneity between these chromosomes is revealed. With the objective to analyze this paradox in greater depth, a meta-structure was discovered that overlaps all 24 human chromosomes. It is based on a set of strong numerical constraints based particularly on Pi, Phi and integer numbers such as 2, 3 etc. A functionality of this fine-tuned structure appears: the structure is 90% correlated with the density of genes per chromosome from the Human Genome project. It is 89% correlated with the chromosome's permeability to intrusion by retroviruses like HIV, 94% with CpG density and 62% with SNP inserts/deletes. Finally, a classification network of the 24 human chromosomes was discovered, including one measuring scale, ranging from 1/Phi (chromosome 4) to 1/Phi + 1/Pi (chromosome 19), which is both correlated with the increasing density of genes and permeability to the insertion of external viruses or vaccines.

3.3.3.6. Unifying all Biological Components of Life: DNA, RNA, Proteins. A powerful basic Pi, Phi based numerical projection law of the C O N H S P bio-atoms average atomic weights was established; Perez [2009a], methods will be published in a forthcoming paper; [Perez [2012]. An integer-based code unifies the 3 worlds of genetic information: DNA, RNA and Protein-aggregating amino acids. Correlating, synchronizing and matching Genomics/Proteomics global patterned images in all coding/noncoding DNA sequences, all biologic data is unified from bio-atoms to genes, proteins and genomes. This code applies to the whole sequence of human genome, produces generalized discrete waveforms. In the case of the whole double-stranded human genome DNA, the mappings of these waves fully correlate with the well-known Karyotype alternate dark/grey/light bands. This "unification of all biological components" is illustrated in Panels 3-4 of Fig. 5; Perez [1988a]. A complete proof of self-similarity within the whole human genome is provided by Perez [2008]. In this "binary code" which emerges from whole human DNA, the ratio between both bistable states is exactly equal to "2" (the space between two successive octaves in music). As shown in Perez [2008] the Top State is exactly matching with a Golden ratio, the Bottom State is also related to the Golden ratio. If PHI = 1.618, it is the Golden ratio, and is phi = 0.618 = 1/PHI, then the "Top" level = phi = 1 / PHI and the "Bottom" level = phi/2 = 1 / 2 PHI. Top / Bottom = 2.

3.3.4. Neural Net Elements are Fractal: Purkinje Neuron Fractal Model

About the same time as the Grosberg-school of thought devoted itself to the analysis of fractal folding of DNA, the School of Recursive Function developed a fractal structural model of a dendritic arborization; Pellionisz [1989]. The seminal concept of "recursion" to the DNA to build a fractal neuron is explicitly argued in point 3.1.3 of that paper: *"Neural Growth: Structural Manifestation of Repeated Access to Genetic Code": "One of the most basic, but in all likelihood rather remote, implication of the emerging fractal neural modeling is that it corroborates a spatial 'code-repetition' of the growth process with the repetitive access to genetic code. This conceptual link between the two meta-geometries of double helix and 'fractal seed' may ultimately lead to precisely pinpointing those exact differences in the 'genetic' code that lead to a differentiation to Purkinje-, pyramidal cell, Golgi-cell or other type of specific neurons. It must be emphasized, however, that establishing a rigorous relation of these 'code sequences' to the genetic code that underlies the morphogenesis of differentiated neurons may be far in the future"*

3.3.5. The Genome is Fractal! Proof of Concept and the Basis of Generalization: Whole Genome Analysis Reveals Repetitive Motifs Conforming to the Zipf-Mandelbrot Parabolic Fractal Distribution Law of the Frequency / Ranking Diagram

This chapter decidedly expands on this point to provide support to the generalization, to further detailing a study heralded earlier on the fractality of a whole DNA; Pellionisz [2006], Simons and Pellionisz [2006b], Pellionisz [2009a].

With a rapidly increasing number of species in which the whole genome is sequenced and DNA is fully available moreover "motif discovery methods" are increasingly available. See the TEIRESIAS algorithm by Rigoutsos and Floratos [1998], the MEME and MAST algorithms by Bailey at al. [1998], and GEMODA algorithm by Jensen et al. [2006], and Kyle et al, [2006], repetitive "motifs" lend themselves as natural units serving as "words". This raises not only the necessity, but a possibility to re-visit the original Zipf law analysis; Flam [1994] and Mantegna et al [1994].

In the study reported here, the recently found short, repetitive sequences ("Pyknon"-s) described by Rigoutsos et al. [2006] are used as more natural "words" than completely arbitrarily picked 3-8 nucleotide sequences. In the human DNA, they found about 128,000 short, repetitive sequence elements, apparently indiscriminately distributed over coding- as well as non-coding regions of the DNA.

Therefore, there is no need, indeed no basis to separate “words” occurring in the DNA either in the regions of “genes” or what used to be called “junk” DNA. In addition, the short, repetitive sequence motifs mined by the TEIRESIAS algorithm, Rigoutsos and Floratos [1998] showed no apparent difference in occurrence either in the “coding” or “non-coding” regions.

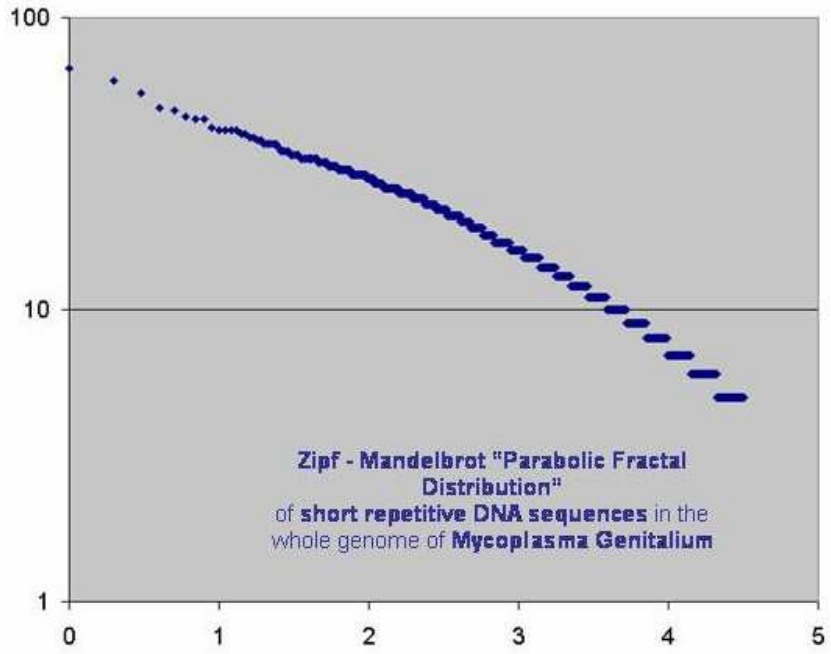

Fig. 6. Zipf-Mandelbrot Parabolic Fractal Distribution curve of short repetitive DNA sequences in the whole genome of Mycoplasma genitalium. *Frequency as a function of rank is parabolic on a log/log scale, after Pellionisz [2009a]. See detailed explanation there and in the text below.*

Using the web-interface by the Group of Rigoutsos at IBM Watson Research Center http://cbcsrv.watson.ibm.com/Tspd.html a “pyknon-type” motif discovery was made for the whole genome of the Mycoplasma Genitalium; the smallest DNA known, Fraser et al [1995].

The web-interface returned the list of short, repetitive DNA sequences in the order of their ranking (as integers) with the frequency of occurrence (also as integers). Results immediately lend themselves to a log-log plotting of the frequency (y) against ranking (x), as seen in the graph below.

Fig. 6 shows the frequency (y) plotted against ranking (x) of “Pyknon-Like-Elements” short repetitive sequences (PLE-s) of the whole DNA of Mycoplasma Genitalium. Results reveal a “Zipf-Mandelbrot Parabolic Fractal Distribution”. Both frequencies and occurrences are shown on a log-log scale. Note that the actual distribution is distinctly different from the linear Zipf’ law. More detailed analysis of the above results reveals by standard curve-fitting that the data-can be modeled by the generalization of Zipf’s Law, defined as the Zipf-Mandelbrot Parabolic Fractal Distribution. The Zipf-Mandelbrot function is given by

$$f(k;N,q,s) = \frac{1/(k+q)^s}{H_{N,q,s}}$$

where $H_{N,q,s}$ is given by

$$H_{N,q,s} = \sum_{i=1}^{N} \frac{1}{(i+q)^s}$$

this may be thought of as a generalization of a harmonic number. In the limit as N approaches infinity, this becomes the Hurwitz zeta function $\zeta(q,s)$. For finite N and $q = 0$ the Zipf-Mandelbrot law becomes Zipf's law. For infinite N and $q = 0$ it becomes a Zeta distribution.

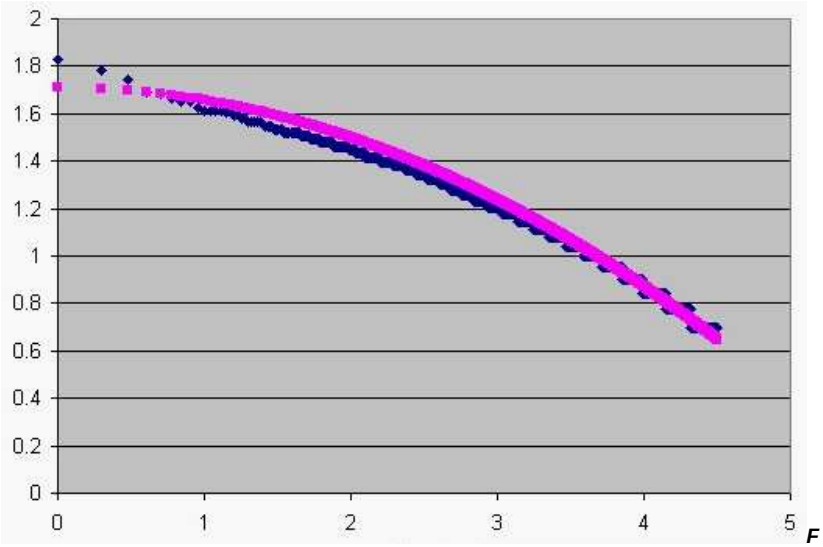

Fig. 7. Curve-fitting (in purple) of the frequency (y) against ranking (x) of "Pyknon-type" short repetitive sequences of the whole DNA of Mycoplasma Genitalium (in blue). The curve reveals a Zipf-Mandelbrot Parabolic Fractal Distribution that can be approximated by the quadratic polynomial of $y = -0.052x^2 - 0.0015x+1.71$*, after Pellionisz [2009b], see detailed explanation there and in the text below.*

In the Parabolic Fractal Distribution, the logarithm of the frequency or size of entities in a population is a quadratic polynomial of the logarithm of the rank; standard curve-fitting approximates the data with the quadratic polynomial $y = -0.052x^2 - 0.0015x+1.71$

As in typical cases, there is a so-called King effect where the highest-ranked item(s) tend to exhibit a significantly greater frequency or size than the model predicts on the basis of the other items.

Data by the Rigoutsos et al. (2006) motif discovery reveal the Zipf-Mandelbrot Parabolic Fractal Distribution curve of frequency against ranking of short repetitive sequences in the entire genome (full DNA) of a free-living organism. It is noteworthy that for the analysis no distinction between the "protein coding" and "non-protein-coding" DNA segments need to be made.

Nonetheless, one might argue that since in the DNA of the Mycoplasma Genitalium contains only <8% "non-coding DNA" (the rest is almost a "wall-to-wall" protein-coding sequence), the found Parabolic Fractal Distribution might be characteristic for the coding DNA.

Therefore, utilizing recent results of identification of "FractoGem"-s (group of "FractoSet"-s, each composed of pyknon-type repetitive short sequences that are found strictly in the non-coding intronic regions of the Presenilin gene of Alzheimer's, data from http://www.fractogem.com), a comparative graph is provided below, applying strictly for non-coding short repetitive sequences; Fig. 8.

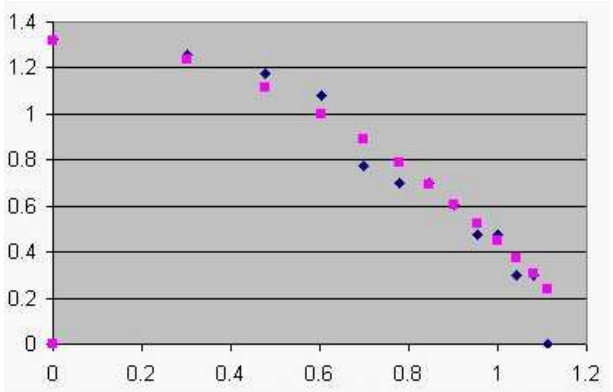

Fig. 8. Zipf-Mandelbrot Parabolic Fractal Distribution curve in the strictly non-coding DNA. *Frequency (y) against ranking (x) data-points are in blue for repetitive short sequences of the FractoGem of the Presenilin intronic areas in Alzheimer's. The curve dotted in purple reveals a Zipf-Mandelbrot Parabolic Fractal Distribution that can be approximated by the quadratic polynomial of $y = -0.85x^2 - 0.022x+1.32$, after Pellionisz [2009b], see detailed explanation there and in the text below*

While the number of data-points are limited since the FractoGem of the human Presenilin intronic areas contains 27 FractoSet formations only and each with a maximal number of 13 "pyknon-type" short repetitive sequences, the Zipf-Mandelbrot Parabolic Fractal Distribution curve appears applicable. Results of this paper may need to be reproduced and extended to the whole genome of species

in comparative genomics to DNA other than that of Mycoplasma Genitalium. Other points of interest are if some coding DNA corpus larger than that of the intronic sequence of a single gene (Presenilin) will yield similar indiscrimination for the Parabolic Fractal Distribution, and if "motif discovery algorithms" other than Tereisias by Rigoutsos and Fluertes (1989) confirm the present study.

Overall, if a "mathematical language" is suspected to be hidden in the DNA (in coding as well as non-coding regions), the thesis of this paper is that currently best candidates for "words" are the short, repetitive segments as revealed by the Tereisias motif discovery algorithm, and the most likely mathematical language is modeled by the Zipf-Mandelbrot Parabolic Fractal Distribution curve of frequency over ranking, a log-log scale.

4. Conclusions

Main conclusions of this chapter are:

- The One-to-One Arrow-model of e.g. "three-neuron-reflex-arc" by Lorente de No [1933] lost to the "All-to-All" matrix-model; Szentágothai [1949] *"Elementary reflex arcs are convenient abstractions rather than real functional units of the nervous system"*).
- The Arrow-model of Genomics (Crick's Central Dogma [1956, 1970]) were obsolete before its birth by the importance of "feedback" by Cybernetics, Wiener [1948], and is superseded by The Principle of Recursive Genome Function; Pellionisz [2008a].
- The massively parallel systems of Neural Nets and Recursive Genome Function are to be mathematically described by multi-component entities (vectors including dual representation), rather than by serial loops.
- Biological System Theory is compelled to identify the mathematics of the system, in a manner to conclude in software enabling algorithms.
- Coordination by the cerebellum is to be characterized by generalized coordinates as in non-Euclidean tensor and fractal geometry.

4.1. Neuronal and Genomic Systems are Governed by Recursive Algorithms of Massively Parallel Networks, not only including, but Surpassing Serial Feedback

The above main conclusions are comprised into the single above statement. The consequences are the following:

- The cerebellar Purkinje Neuron is fractal, similarly the folding of the DNA is fractal.
- The Zipf-Mandelbrot Parabolic Fractal Distribution curve of the full DNA of an organism clinches that the Genome is Fractal.
- Computational Unification is made possible by the full utilization of recursion, deploying Neural Net Algorithms.
- Neural Nets are applicable because the Recursive Genome Function massively parallel.

4.2. Application of Fractal Genomics is Already Here

While to most people "fractals" are either pretty pictures or some exotic branch of mathematics, as usual in the history of mathematics, practical applications already exist.

4.2.1. Friedreich Spinocerebellar Ataxia

Since the function of the cerebellum is sensorimotor coordination (by acting as a metric tensor), symptoms of aberrant cerebellar function is often called "dysmetria" (literally meaning that the precise metric is absent). Research of the great number of varieties of "ataxia" such as the as lack of proper cerebellar coordination is a very large, active field, as reviewed recently by Manto and Marmolino [2009].

A specific kind of dysmetric cerebellar disorder is the Friedreich's Spinocerebellar Ataxia, see extensive reviews on Friedreich Ataxia by Timchenko and Caskey (1999), Pandolfo [2009]. This autosomal recessive congenital disease is known to be caused by a GAA triplet "run" in the first intron of the FXN (originally, known as X25) gene on 9q13-q21 that codes for a protein frataxin. This protein is essential for mitochondria, as in its absence iron builds up and causes free radical damage in nerve cells (such as in the cerebellum) and in muscle cells – that is often the cause of heart failure in those affected by Friedreich. It is a particularly interesting case, since the GAA "run" is intronic, thus it does not result in the production of abnormal frataxin proteins. Instead, the mutation in the regulatory sequence causes gene silencing; Castaldo et al [2008]. Thus, an insufficient amount of Frataxin – or in more serious cases a long tract of GAA repeats, structurally weakens the DNA strand and the chromosome through breakage, as evidenced through *in vivo* yeast studies. While a characteristically genomic disease, Friedreich Ataxia is on the verge of therapy; Marmolino and Acquaviva [2009].

For the reasons above as reported earlier Pellionisz [2009a] a structural analysis of fractal defects was performed using FractoSoft Miner of HolGenTech, Inc. As shown in Fig. 9 the fractal defect disrupting regulatory function was found below the GAA triplet repeat in the middle of an (intronic) Alu repeat (see PLE-s displayed in various colors).

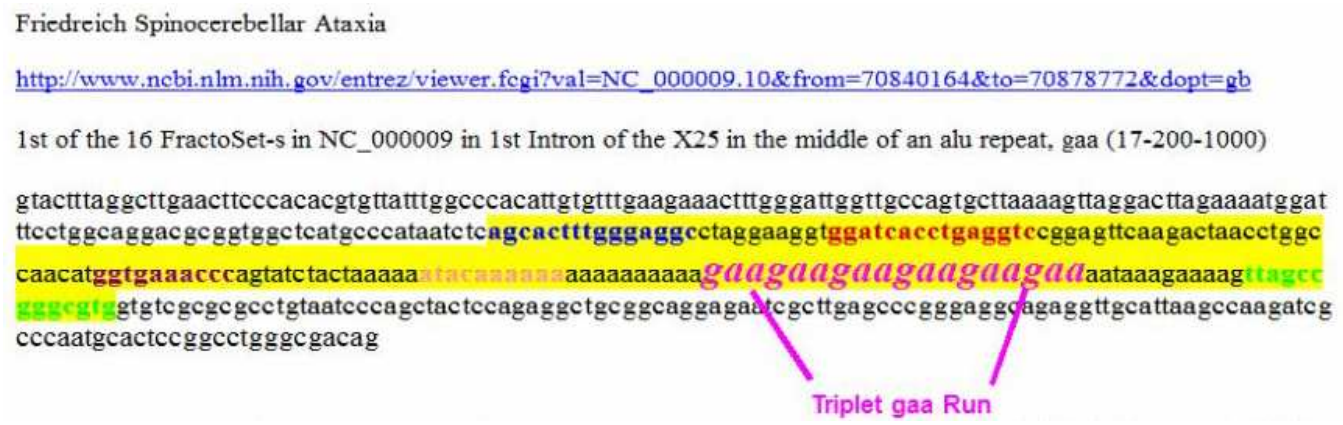

Fig. 9. Friedreich Spinocerebellar Ataxia is known to be caused by a GAA triplet repeat at a known locus. Fractal analysis reveals a FractoSet of Pyknon-like elements (short oligos shown in different colors). *It is conspicuous that the fractal defect is disrupted by the GAA triplet repeat, after Pellionisz [2009b], see detailed explanation there and also in the text below.*

Examination of long (or full) DNA sequences for fractal defects is made important by the logic that since the genome is fractal, the actual sequence must obey the fractal laws for proper function. For about a dozen hereditary conditions such fractal defects have been identified. This is promising also for a very important practical-logistical reason. Our rapidly increasing tally of full DNA sequences shows "structural variants" how the individual genomes are different from one-another. The different bases can be counted by millions. Therefore, a mere cataloguing of such variants is unlikely to be a solid strategy of hunting down diseases. Some variants most likely only cause "human diversity". Perhaps only a much smaller set of variants could be the root causes for diseases. Mathematically speaking, in the most famous fractal, the Mandelbrot-set [1983], the mind-boggling "complexity" arises from the rather simple equation Z=Z^2+C. In the equation C is a constant, that may have the value of c or D (etc.) and the fractal set still emerges; just looks somewhat different. The differences between individual genomes, therefore, fall into two separate classes. "Structural variants" can be neatly parsed into what we call "parametric structural variants" (PSV-s, e.g. various values of the constant in the equation). The c, or D will not violate the pristine fractal equations. However, human genomes are likely to harbor "syntax structural variants" (SSV-s). These are alterations can render the fractal equation invalid; Z≠Z^2+C, thus the genome's own fractality may be compromised (as with Friedreich Ataxia) or even grossly violated (as with cancers). These "syntax structural variants" SSV-s can be mathematically expected to be direct causes of genome misregulation.

Using a computer code metaphor to illuminate the above argument, an algorithm can be implemented with harmless "structural variants of lines of code". In these cases the versions of the code would all run, but perhaps some versions of the code would more rapidly or slowly converge than others. However, if some lines of code would contain syntax-errors, the code not only would never run, but could not even be compiled. Beyond the above proof of concept with Friedreich', the perspective of genomic cancer diagnosis looms, by means of Fractal Defect Mining for SSV-s. This opportunity is further detailed in the "Cancer" section of "Future Directions".

4.2.2. Application of Fractal Genomics for Cancer

Cancer is widely regarded as "the disease of the genome". Scientific results abound stating that that the progression of genome mis-regulation causes massive amounts of structural variants of the DNA; see recent reviews; Meyerson et al. [2011] and Ozery-Fleto [2011].

In cerebellar tumors, it was found that sonic hedgehog signaling regulates the growth and patterning of the cerebellum; Dahmane [1999]. Also, retinoid-related orphan receptors (RORs) were found to play critical roles in cancer, development, immunity, circadian rhythm, and cellular metabolism; Jetten [2009]. A link between RORγ and cancer is emerging from studies showing increased expression of Th17-associated genes, including (an at least 3-component vector); RORγ, IL-17, and IL-23.

A particularly strong study suggests a possible role for RORα in cancer development; Jetten [2009]. *"The ROR α gene spans a 730 kb genomic region that is located in the middle of the common fragile site FRA15A within chromosomal band 15q22.2 .... Common fragile sites are highly unstable genomic regions found in all individuals and are hotspots for deletions and other genetic alterations that may lead to altered expression and function of genes encoded within these regions. Common fragile sites have been implicated in several human diseases and are associated with a number of different cancer types [Smith et al., 2006]. Genomic instability within FRA15A might lead to changes in the expression of RORα and play a role in the development of certain cancers. This hypothesis is consistent with observations showing that RORα mRNA expression is often down-regulated in tumor cell lines and primary cancer samples… Moreover, studies examining gene expression profiles in various cancers identified ROR α as a gene commonly down-regulated in several tumor types, particularly breast and lung cancer…Analysis of the methylation status of a series of genes identified ROR α as one of methylation-silenced genes in gastric cancer cell lines [Yamashita et al., 2006]"*.

The latter is in agreement with the concept that reduced expression of RORα expression positively correlates with tumor formation. A major factor for such dramatic alterations appears to be the mis-regulation due to hypo-methylation of DNA; Hansen et al. [2011].

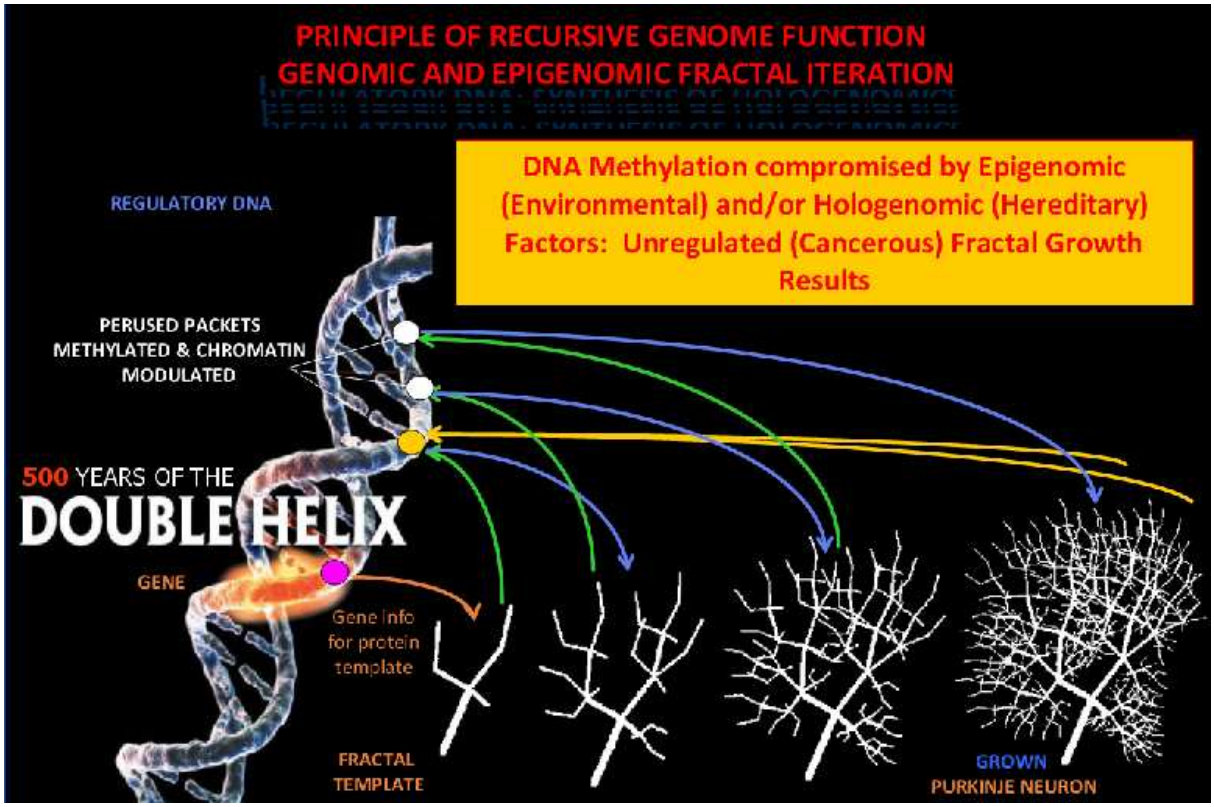

Fig. 10. Genomic and epigenomic fractal iteration derailed. *(From minutes 30:00 of YouTube "Is IT Ready for the Dreaded DNA Data Deluge"?) Further explanation is in the video Pellionisz [2008b] at 30:00 minutes and in the text below.*

In terms of the fractal iterative recursion of multi-genic vectors through matrices, the tentative diagram below illustrates the concept exposed for general audience [Pellionisz 2008b, at minute 30:00]. In Eigenstates perused master switches are methylated (in the diagram of Fig. 10 shown by white "cookie") and DNA-enhancer and suppressor vectors force fractal recursive iteration into next stage of hierarchy. In the diagram, an erroneous methylation (shown by yellow "cookie") would result in perusing a master switch in an uncontrolled manner – thus the fractal growth of the neuron would degenerate into a proliferation, instead of stopping at the full grown state of the cell.

It should be emphasized, that the tentative scheme shown in Fig. 10 is a seminal concept where the diagram greatly simplifies a hypothetical cancerous growth due to hypo-methylation of the genome. First, the diagram shows only a Purkinje neuron, though it is observed by both Dahmane [1999] and Jetten [2009] that cerebellar cancers depend on an interaction of Purkinje- and granule cells. Also, it should be pointed out that the recursive lines between DNA regions and protein-structures represent the action of not a "single-gene to single-RNA, to single-protein" loop, but recursion of multicomponent vectors. For instance, as shown in Jetten [2009] an at least three-component vector (of RORγ, IL-17, and IL-23).

From a flood of evidence it is clear, that development of the cerebellum requires the multi-dimensional co-regulation of vectors of genes; Oberdick [since 1993], Barski et al [2002]. It might take substantial time to assemble a comprehensive map of genic and regulatory sequences that result normal or pathological (e.g. cancerous) cerebellar neural networks.

5. Future Directions

5.1. Theory of Recursive Algorithms

The Principle of Recursive Genome Function peer-reviewed paper Pellionisz [2008a], also disseminated for general audience [Pellionisz 2008b] and presented for debate at Cold Spring Harbor Labs [Pellionisz 2009b] laid out an agenda also in practical terms [Pellionisz 2010] calling for substantial time and resources. As usual with a new set of principles, future directions abound beyond the originally outlined boundaries. It is understood that the tasks for theory-development outlined below will require substantial time, perhaps generations, and sizable resources.

5.1.1. Neural Net Algorithms Comprise Massively Parallel and Coordinated Genome Function

A central thesis of this chapter is that both neuroscience and genomic is characterized by the "many-to-many" concept, that has not been emphasized sufficiently in the past of genomics. Neural Net algorithms, both existing, see e.g. Anderson et

al. [1990] and to be developed are most likely be deployed in the analytics of massively parallel genome function. Much of the gene-expression in "coordinated genome function" of "single genes" e.g. in Operon-theory, Jacob and Monod [1961] was based on a mind frame reminiscent of the "single reflex loop" of early neuroscience and thus coordinated genome function could not gain as much ground for the past half of a Century as it was inherent in their initiative.

5.1.2. Integration of Neural Net and Fractal Algorithms

The principle that coordinated genome function is based on recursion puts enormous emphasis on the accelerated development of the theory of recursive algorithms suitable for an algorithmic (software-enabling) understanding of coordinated genome function. In this regard, both Neural Net algorithms as well as Fractal Geometry algorithms have to be much further developed and integrated. While Fractal Iterative Recursion is already featured, this chapter emphasizes that the recursion is not a "single loop" but is implemented in a massively parallel manner. Moreover, the genome is certainly not monofractal, but multifractal, thus algorithmic development must be directed accordingly, Barnsley [2006].

5.1.3. Develop and Integrate Quantum Theory of Neuroscience and Genomics

Neural firing of spikes and the A, C, T and G bases of the DNA will similarly require a recognition that science is facing a quantum system both in neuroscience and genomics. The "aperiodical" covalent bindings predicted by the seminal idea of Schrödinger [1944] preceded the discovery of DNA bases that establish such bindings – perhaps a reason why emergence of quantum theory is sluggish compared e.g. to that of physics. It is a question, however, if the discrete units are A, C, T and G bases, or, rather the quanta are codons (both amino-acid coding, as well as "pervasively transcribed" non-coding triplets), short repetitive motifs (but certainly not arbitrarily picked 3-8 character "words"), or fractal "pyknon-like elements" (PLE-s).

In a theoretical unification the question will arise if in Recursive Genome Function: Contravariant (Amino-Acid producing genic vectors) and Covariant (Protein-bonding DNA-site vectors) - converge and thus obey the Fractal Weyl's Law on Fractal Quantum Eigenstate*;* Shepelyansky DL [2008], see the original Weyl Law [1912].

5.2. Public Domain Agenda in Industrialization of Genomics: Local and Global Fractal Dimension as a Standard Definition for Optimally Distinguishing Cancerous and Control Genomes Based on their Abstract Measure of "Roughness"

Given the Conclusion that "the genome is fractal" there is an immediate need, to accomplish by a common and publicly available standard, worked out by the joint effort of all concerned (genome informatics firms, cancer- and genome centers, etc.). The goal is to arrive at a commonly accepted best performing definition of the global and local "abstract roughness" (fractal dimension), in a manner optimized for detection of misregulated (cancerous) genomes by bringing out the difference in fractality of cancerous and control DNA.

Genomics and the "New War on Cancer"; Watson [2008, 2009] could greatly benefit from a common focused effort of leading mathematically minded genomists devoted to this vital practical problem of postmodern genomics.

Fractal dimension of physical objects, normally in two- or three-dimensional spaces can follow the definition based on how fully the object fills the available space. For instance, the Hilbert-curve, shown on the Science cover –article by Erez-Lieberman et al. [2009] elaborating on the seminal concept of Grosberg et al, 1988, 1993] shows the fractal folding of DNA – squeezing a 2m long double helix into the 6 micron diameter nucleus of a cell; where the Hilbert-curve is not only "knot-free" in order to ensure uninterrupted transcription, but is also ultra-dense, i.e. "space-filling" with the physical fractal dimension of 3.

It needs to be pointed out, that "fractal dimension" can be defined not only for actual physical objects, but the "roughness" of e.g. the double helix (say, if you would run through the thread your fingers equipped to feel, like a brail pattern, the A, C, T and G bases separately) can also be measured – given that both the "abstract object" and the "abstract embedding space" is appropriately defined.

In the past, there were several attempts at defining "DNA fractal dimension". Berthelsen et al. [1992, see their Fig. 4] used both a 2-dimensional embedding in a space spun by AT horizontal- and CG vertical axes, as well as a 4-dimensional embedding in a space spun by the AA:TT horizontal, CC:GG vertical, AG:GA-GT:TG and AC:CA-GT:TG diagonal axes. The Grosberg-school of fractal DNA, beyond their seminal concept of fractal folding of DNA; Grosberg et al, 1988, 1993] also revisited the issue of fractality of DNA texts, Borovik et al [1994].

The numerous early DNA fractal dimension studies were triggered by Mandelbrot [1983] but were conducted much before the now multiple supporting facts available that both the genome is fractal (see section "Zipf-Mandelbrot Parabolic Fractal Distribution" of this chapter, the entire double helix folds in a fractal manner Erez-Lieberman [2009], brain cells such as the Purkinje neurons are fractal; Pellionisz [1989] – plus our novel explosive set of data that not only the surface of cancerous cells differs in spatial fractal dimension from the control cells; Dokukin et al, [2011], but rather, there is a massive re-arrangement in the structure (obviously affecting the local and global "roughness") of cancerous genomes.

Given the amount of rapidly amassed data of cancerous and control full human DNA, it is an urgent as well as eminently feasible project is to arrive at the definition of both “the abstract DNA roughness” as well as the “abstract space in which it is embedded” with the definitions optimized for distinguishing cancerous genomes from their pristine (control) sequences.

In fractal theory, objects can be measured by different standards (“yardsticks”). The famous question *"How Long Is the Coast of Britain?"* by Mandelbrot [1967] can be answered in an infinite number of ways - as the length minimal or infinite – depending on how science defines the "yardstick" with which the same object is to be measured. Likewise, in defining the abstraction of global and local “roughness” of the genome, appropriately embedded into an abstract multidimensional space, it is reasonable to expect that cancerous deterioration can be tracked by “measurement of local and global fractal dimension”, thus providing a diagnostic tool – before unregulated/malformed proteins appear as the result of genomic rearrangements.

By what yardstick does Industrialization of Genomics (starting with present R&D of Cancer) can best measure the fractal difference characteristic to cancerous DNA (fragments)? While “fractal dimension” mathematical literature is rich, genomic/methylomic data is only presently available to identify the most suitable mathematical definition for this novel but life-or-death application.

While earlier attempts focused on A, C, T and G bases to define an abstraction (embedded either into a two-dimensional, or four dimensional abstract space, spun over nucleotides), novel research points into the possibility of defining an abstract space of codons; Perez [2011], wherein both “protein-coding codons” and the “pervasively transcribed” so-called “non-coding triplets” would also be embedded. Further considerations include methylation and chromatin modulation – rendering segments of DNA “unreadable” temporarily or permanently. It is a matter of definition of an unreadable (silent) DNA segment is totally smooth (with fractal dimension zero) – or to the contrary, like an unpaved terrain, “infinitely rough”, thus impossible to be traveled. Another matter of definition is in what abstract space are the abstract objects embedded. In codon-space embedding, or pyknon-space embedding the measures are not only numerically different, but they are likely to bring out the differences in fractality of cancerous and control DNA more-or-less revealing.

Presently there is enough public DNA (with control) of cancerous sequences, with already plenty of evidence for massive pathological alterations. It is a task for a community of leading experts to work out by what definitions we could get the best standard to spot the fractal genomic alteration associated with the progression of the disease*. "The Fractal Yardstick for Cancer"* will emerge as a public domain accomplishment, yielding an optimized and standard definition for genome analytics.

Public domain DNA data are to be downloaded from Cancer Centers, worldwide. With the body of fractal literature reviewed, box-counting and other available algorithms will be critically applied to provide the best practical definition to bring out differences in terms of the fractal dimension of DNA (entire or fragmental). It is of particular significance that formerly DNA fractal dimension was not focusing on the methylation of bases, though by rendering certain sequences unreadable the fractal dimension of the retrievable DNA information is most certainly altered. A community effort also provides the opportunity of running and re-running benchmark tests as the work of the study-group develops by cloud computing on the same body of data. The initiative plans for deploying not only “public clouds” (composed of serial computers) – but because of considerations of human data privacy (HIPPA), later deployment of proprietary algorithms e.g. by fiercely competitive Big Pharma (about as unlikely to rely entirely on open-source of algos and software as financial computing retains proprietary), this initiative proposes simultaneous deployment of “private clouds” – composed of hybrid computers for speed and physical efficacy (footprint and energy conservation). A genome informatics specialist with cross-disciplinary experience might be welcomed to lead this initiative. The group of top experts is expected to define the mathematical- and computing strategy and weaponry of the New War on Cancer.

5.3 Proprietary Agenda in Industrialization of Genomics

The “Battelle Report” [2011] sized up this May the Economic Impact of the Human Genome Project in how $3.8 Bn investment drove $796 Bn in economic impact, created 310,000 jobs and launched the genomic revolution. Not unlike how the development of the science of nuclear physics was a necessary but unsatisfactory condition to develop nuclear industry, genome informatics should be mindful that the industrialization of genomics might at any time become unsustainable unless the scientific challenge of understanding coordinated genome function in an algorithmic software-enabling manner is met by an accelerated agenda. The scientific challenge is complicated by the very beneficial involvement of the private sector (global informatics and product companies, like Samsung, Procter & Gamble, Nestlé, Unilever, and of course global Pharma companies, like Genentech/Roche, as well as private hospital systems with Cancer Centers in the lead). Given the fact that there are about 1,000 Cancer Centers in the USA alone, and over 400 cancer drugs involving practically all Pharma companies, as well as the computerization of both the hardware and software of hospital systems being a lucrative business, Industrialization of Genomics is likely to follow previous complex models. Most notably, those of defense, financial computer science and industry – with intertwining public and fiercely competitive thus strictly proprietary intellectual property, based on in-house science.

5.3.1. Hybrid Computation on Private Clouds

As assessed recently, Schadt et al. [2010], Industrialization of Genomics enables individual laboratories to affordably generate terabytes or even petabytes of data. Fortunately, as pointed out in a general presentation Pellionisz [2008b] “Is IT ready for the Dreaded DNA Data Deluge,” the real challenge is not the readiness of information technology, since earlier data-intensive applications (defense-, nuclear-, financial-, meteorological-, graphic-computing, etc.) have all been dealt with the immense computing industry. Thus, the main challenges are in Information Theory as Genome Informatics is applied towards an algorithmic understanding of genome-epigenome (hologenome) regulation. Some of the scientific algorithms, just as in financial computing, are fiercely proprietary

(not only to provide accurate predictions, but deliver them faster than the competition). Industrialization of Genomics emerges with entire segments (biodefense, private-domain wellness and health-care) in a proprietary fashion. Additionally, since genomics deals with human data that are legally mandated (in the USA, by HIPPA) to be handled in a confidential manner, not only algorithm-security, but data-security is also indispensable. Genome Computing Architecture, therefore, emerges with special needs and solutions; Pellionisz [2009b].

Thus, though global IT firms (Microsoft, Amazon, Google, Facebook, etc.) have mastered handling petabytes by computing architectures distributed over massively parallel systems, only the transient research and development phase of the Industrialization of Genomics, when many volunteers forgo privacy for the interest of faster progress, will permit the standard "public cloud computing." Though public clouds are increasingly more secure e.g. by encryption, Baylor at Houston, Texas already decided that for genome computing the appropriate solution is a "private cloud," moreover a closed system that is composed by the hybrid (serial/parallel) computers. These platforms, available off the shelf from many companies with a highly successful record of applying them in defense-financial- etc. computing, additionally provide e.g. for hospitals the small footprint, low energy consumption advantages, and most of all the speed that will be required for hospital applications, when a biopsy tissue-sample will have to be sent to the local computing lab and results could be relayed back to the operating theater while the patient is still on the table. Sequencing and Analytics all performed locally, fast and affordably, without shipping hard-disks or uploading and downloading data.

5.3.2. Consumer Genomics in Continuous Customer Care

Industrialization of Genomics started with Consumer Genomics that the FDA of the US does not regulate, 23andMe, Inc. and Navigenics, Inc. While the market of health care (genomic diagnosis, pharmaco-genomics of patients) is likely to be restrained by regulatory industries, there is already a global trend, both in Europe and Asia, to extend the benefits of genome interrogation, genome sequencing and genome analytics to vast masses of consumers; Pellionisz [2010] "Shop for Your Life - HolGenTech at PMWC2010". This trend will switch one-time analytics into continuous customer care, known in business as the most lucrative "repeat customer mode". The announcement of Samsung, starting to provide genome analytics by September 1st, 2011 signaled a change of times. A Genome-Based Economy has already commenced.

6. Acknowledgements

Upon presentation, an advice was received from Dr. Hamilton O' Smith (Venter Institute), to run the Zipf-Mandelbrot Fractal Parabolic Distribution Curve-test, as a control, with an identical number of A, C, T and G-s, randomly generated. Not only the control-result did not show the Curve, but generated zero repetition for the identical overall length and motif-requirements (there were no points to compose any curve). The creative suggestion is gratefully acknowledged. The authors thank Paul Shapshak PhD, Division of Infectious Disease and International Medicine and Dept. of Psychiatry and Behavioral Medicine, University of South Florida College of Medicine, Tampa, Florida, to help prepare the manuscript and to Prof. Sergey Petoukhov, Academician, Moscow, for reference to Gazalé and appreciative comments of the Chapter on dual valence, the RNA system serving as a Genomic Cerebellum".

7. References

1. Amari S (1991) Dualistic Geometry of the Manifold of Higher-Order Neurons. Neural Networks 4(4):443-451
2. Anderson JA (1990) 351-355 In: Anderson JA, Pellionisz A, Rosenfeld E. Neurocomputing II. Directions of Research. MIT Press, http://usa-siliconvalley.com/inst/pellionisz/90_anderson/90_anderson.html
3. Anderson JA, Pellionisz A, Rosenfeld E (1990) Neurocomputing 2. Directions of Research. MIT Press http://books.google.com/books/about/Neurocomputing.html?id=kk6NWGht1UIC
4. Arneth BM (2010) Sequence variability and sequence evolution: An explanation of molecular polymorphisms and why many molecular structures can be preserved although they are not predominant. DNA Cell Biol. 29(10) Pp. 571-576. DOI: 10.1089/dna.2009.0942
5. Bailey, Timothy L, Gribskov M (1998) Methods and statistics for combining match scores (MEME, MAST). J of Comput. Biol. 5:(211-221)
6. Barnsley MF (2006) Superfractals. Cambridge Univ. Press http://www.amazon.com/SuperFractals-Michael-Fielding-Barnsley/dp/0521844932/ref=sr_1_1?ie=UTF8&qid=1320384324&sr=8-1
7. Barski JJ, Lauth M, Meyer M (2002) Genetic targeting of cerebellar Purkinje cells: History, current status and novel strategies. The Cerebellum:1:111-118 http://www.ncbi.nlm.nih.gov/pubmed/12882360
8. Bonhoeffer S, Herz AV, Boerlijst MC, Nee S, Nowak MA, May RM (1997) No signs of hidden language in noncoding DNA. Phys. Rev. Lett. 76(11)
9. Battelle Technology Partnership (2011) Economic Impact of the Human Genome Project. How a $3.8 Bn investment drove $796 billion in economic impact, created 310,000 jobs and launched the genomic revolution http://www.battelle.org/publications/humangenomeproject.pdf
10. Bertalanffy L von (1934) Untersuchungen über die Gesetzlichkeit des Wachstums. I. Allgemeine Grundlagen der Theorie; mathematische und physiologische Gesetzlichkeiten des Wachstums bei Wassertieren. Arch. Entwicklungsmech.131:613-652
11. Berthelsen CL, Glazier JA, Skolnick MH (1992) Global fractal dimension of human DNA sequences treated as pseudorandom walks. Phys. Rev. 45(12) http://biocomplexity.indiana.edu/jglazier/docs/papers/20_DNA_Analysis.pdf
12. Bieberich E (1999) Structure in human consciousness: Fractal approach to the topology of the self perceiving an outer world in an inner space. http://cogprints.org
13. Bieberich E (2011) Introduction to the Fractality Principle of Consciousness and the Sentyon Postulate. Cogn Comput. DOI 10.1007/s12559-011-9104-5 http://www.springerlink.com/content/964347x7251k0p36/
14. Bloedel JR, Tillery SI, Pellionisz AJ (1988) Experimental-theoretical analysis of the intrinsic geometry of limb movements. Neurosci. Abst. 14:952
15. Borovik AS, Grosberg AY, Frank-Kamenetskii MD (1994) Fractality of DNA texts. J of Biomol. Structure and Dynamics. 12(3):655-669
16. Braitenberg V (1967) Is the cerebellar cortex a biological clock in the millisecond range? Progr Brain Res. 25:334-346
17. Cartieri FJ (2009) Darwinism and Lamarckism before and after Weisman: A historical, philosophical, and methodological analysis. Univ. Pittsburg; 1-54. http://etd.library.pitt.edu/ETD/available/etd-05152009-110207/unrestricted/CartieriFJ_BPhil_ETD2009.pdf
18. Castaldo I, Pinelli M, Monticelli A, Acquaviva F, Giacchetti M, Filla A, et al (2008) DNA methylation in intron 1 of the frataxin gene is related to GAA repeat length and age of onset in Friedreich ataxia patients. J Med Genet. 45(12):808-12.
19. Chatzidimitriou-Dreismann CA, Steffer RM, Larhammar D (1996) Lack of biological significance in the "linguistic features" of noncoding DNA - a quantitative analysis. Nucleic Acids Res; 24(9):1676-81
20. Church GM (2005) The personal genome project. EMBO and Nature Pub. Group, Mol Syst Biol. 1:30 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1681452/?tool=pmcentrez
21. Chiappelli F, Shapshak P, Commins D, Singer E, Minagar E, Oluwadara O, et al. (2008) Molecular epigenetics, chromatin, and NeuroAIDS/HIV: Immunopathological implications. Bioinformation. 3(1): 47–52 http://www.usa-siliconvalley.com/chiappelli.pdf

22. Churchland PS (1986) Neurophilosophy: toward a unified science of the mind-brain. MIT Press http://usa-siliconvalley.com/inst/pellionisz/85_churchland/85_churchland.html
23. Collins F (2007) New findings challenge established views on Human Genome, Nature, June 14, 2007; http://www.genome.gov/25521554
24. Crick F (1970, seminal notion 1956) Central Dogma of Molecular Biology. Nature 227(8):561-563 http://profiles.nlm.nih.gov/ps/access/SCBCCH.pdf 1956: see http://www.junkdna.com/
25. Dahmane N, Ruiz i Altaba A (1999) Sonic hedgehog regulates the growth and patterning of the cerebellum. Development 126, 3089-3100 http://dev.biologists.org/content/126/14/3089.full.pdf
26. Descartes R (1629) Treatise of man. Prometheus Books (in English, 2003) http://www.amazon.com/s/ref=nb_sb_noss?url=search-alias%3Daps&field-keywords=Treatise+of+man&x=12&y=16
27. Daunicht W, Pellionisz A (1987) Spatial arrangement of the vestibular and the oculomotor system in the rat. Brain Res. 435:48-56. http://www.ncbi.nlm.nih.gov/pubmed/3501327?dopt=Abstract
28. Dokukin ME, Guz NV, Gaikwad RM, Woodworth CD, Sokolov I, (2011) Cell surface as fractal: Normal and cancerous cervical cells demonstrate different fractal behavior of surface adhesion maps at the nanoscale, Phys. Rev. Lett. 107:028101
29. Dupré J, Barnes SB (2008) Genomes and what to make of them. Univ. of Chicago Press
30. Dow RS, Moruzzi G (1958) The physiology and pathology of the cerebellum. Univ.Minnesota Press
31. D’Angelo E, Mazzarello P, Prestori F, Mapelli J, Solinas S, Lombardo P, Cesana E, Gandolfi D, Congi L (2010) The cerebellar network: From structure to function and dynamics. Brain Res. Rev. 5-15 Published by Elsevier
32. Elnitski L, Piontkivska H, Welch JR (2011) Advances in genomic sequence analysis and pattern discovery; Science, Engineering and Biology Informatics, Vol. 7 World Scientific Publishing, Chapter 3. by Fedorov A, Fedorava L; 65-93
33. Erez Lieberman-Aiden et al. (2009) The comprehensive mapping of long-range interactions reveals folding principles of the Human Genome. Science 326. DOI: 10.1126/science.1181369
34. Eccles J, Ito M, Szentágothai J (1967) The cerebellum as a neural machine. Springer Verlag
35. Eckmiller R (1990) Advanced neural computers. North Holland http://www.amazon.com/Advanced-Neural-Computers-R-Eckmiller/dp/0444884009/ref=sr_1_2?ie=UTF8&qid=1319745602&sr=8-2
36. Finger S (1994) Origins of neuroscience: a history of explorations into brain function, Oxford Univ. Press
37. Fiori S (2008) Lie-group-type neural system learning by manifold retractions. Neural Networks. 21(10):1524-1529 http://202.114.89.42/resource/pdf/2174.pdf
38. Fire A, Xu S, Montgomery M, Kostas S, Driver S, Mello C (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391 (6669): 806–11. doi:10.1038/35888. PMID 9486653
39. Flam F (1994) Hints of a language in junk DNA. Science 266:1320
40. Flourens MJP (1824) Recherces expérimentalses fur les propriétés et les sonctions du systéme nerveux dans les anomaux vertébres. Crevot, Paris
41. Foucher I, Montesinos ML, Volovitch M, Prochiantz A, Trembleau A (2003) Joint regulation of the MAP1B promoter by HNF3β/Foxa2 and Engrailed is the result of a highly conserved mechanism for direct interaction of homeoproteins and Fox transcription factors. Development 130, 1867-187 doi:10.1242/dev.00414 http://dev.biologists.org/content/130/9/1867.full.pdf
42. Francis J (2008) Philosophy of mathematics, Global Vision Publishing House
43. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD et al. (1955) The minimal gene complement of Mycoplasma genitalium. Science.;270(5235):397-403 NCBI ascension number NCBI L43927
44. Gardner M (1970) Mathematical games. The fantastic combinations of John Conway’s new solitaire game “life”. Scientific American 223(Oct):120-123 http://ddi.cs.uni-potsdam.de/HyFISCH/Produzieren/lis_projekt/proj_gamelife/ConwayScientificAmerican.htm
45. Gazalé MJ (1999) Gnomonl: from pharaohs to fractals. Princeton Univ. Press http://books.google.com/books?id=R0d76m-Be10C&printsec=copyright&source=gbs_pub_info_s&cad=3#v=onepage&q&f=false
46. Gibbs RA, Jeffrey Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER et al (2007) Evolutionary and biomedical insights from the rhesus macaque genome science 316(5822)222-234 http://www.sciencemag.org/content/316/5822/222.long
47. Gielen CCAM, van Zuylen EJ (1985) Coordination of arm muscles during flexion and supination: Application of the tensor analysis approach. Neuroscience, 17, 527-539 http://www.sciencedirect.com/science/article/pii/030645228690028X
48. Glasner ME, Yen CC, Ekland EH, Bartel DP (2000) Recognition of nucleoside triphosphates during RNA-Catalyzed primer extension. Biochemistry, 39:15556-15562
49. Grosberg AY, Nechaev SK, Shakhnovich EI, (1988) The role of topological constraints in the kinetics of collapse of macromolecules, J. Phys. France 49,;2095-2100 http://hal.archives-ouvertes.fr/docs/00/21/08/91/PDF/ajp-jphys_1988_49_12_2095_0.pdf
50. Grosberg A, Rabin Y, Havlin S, Neer A (1993) Crumpled globule model of the three-dimensional structure of DNA. Europhys. Lett. 23;373-378 http://havlin.biu.ac.il/PS/scan189.pdf
51. Hansen KD (2011) Increased methylation variation in epigenetic domains across cancer types. Nature Genetics pp. 1-8
52. Haussler D (1995) A generalized hidden Markov model for DNA parsing. Extended abstract of talk for the Workshop on Gene-Finding and Gene Structure Prediction, Univ. of Pennsylvania
53. Hood LR (2002) My life and adventures integrating biology and technology. Commemorative lecture given when awarded the 2002 Kyoto Prize in Advanced Technologies
54. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational properties. Proc. Nat. Acad. Sci. (USA) 79:2554-2558
55. Izzo JA, Kim N, Elmetwaly S, Schlick T (2011) RAG: An update to the RNA-As-Graphs resource. BMC Bioinformatics 2011, 12:219 http://www.biomedcentral.com/1471-2105/12/219
56. Jacob F, Monod JJ (1961) Genetic regulatory mechanisms in the synthesis of proteins. Mol. Biol. 3:318-56
57. Jansen J, Brodal A (1954) Aspects of cerebellar anatomy. The Wistar Institute of Anatomy and Biology
58. Jensen KL , Styczynsk MPi, Rigoutsos I, Stephanopoulos GN (2006) A generic motif discovery algorithm for sequential data (GEMODA) Bioinformatics 22(1):21-28
59. Jetten AM (2009) Retinoid-related orphan receptors (RORs): critical roles in development, immunity, circadian rhythm, and cellular metabolism. Nuclear Receptor 7; 1-32, DOI: 10.1621/nrs.07003; http://www.nursa.org/retrieveFile.cfm?type=NRS&file=nrs07003.pdf
60. Kornberg A, Baker TA (1992) DNA replication. Univ. Science Book
61. Kuhn TS, (1962) The structure of scientific revolutions, 1st. ed., Chicago: Univ. of Chicago Press
62. Kyle LJ, Styczynski MP, Rigoutsos I, Stephanopoulos GN (2006) A generic motif discovery algorithm for sequential data. Bioinformatics, 22 (1);21–28 http://bioinformatics.oxfordjournals.org/content/22/1/21.full.pdf
63. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J (et al.) (2001) Initial sequencing and analysis of the human genome. Nature 409:860-921 http://www.nature.com/nature/journal/v409/n6822/full/409860a0.html
64. Laczkó J, Pellionisz AJ Peterson BW. and Buchanan TS (1987) Multidimensional sensorimotor "patterns" arising from a graphics-based tensorial model of the neck-motor system. Soc. Neurosci. Absts. 13., V.1. p. 372.
65. Laczkó J, Pellionisz A, Jongen, H and Gielen CCCM (1988) Computer modeling of human forelimb muscle activation in multidimensional intrinsic coordinate frames. Soc. Neurosci. Absts. pp. 14-2:955
66. Lestienne F, Liverneaux P, Pellionisz A (1988) Morpho-anatomy of sub-occipital muscles in monkey: a tensor model of the musculo-skeletal head-neck system. Reunion Commune de la Physiological Society et de l'Association des Physiologistes, France 2 July 1988. Proc. Physiol. Soc. J. Physiol. (Lond) 2P
67. Llinás R, Roy S (2009) The ‘prediction imperative’ as the basis for self-awareness Phil. Trans. R. Soc. B (2009) 364, 1301–1307 doi:10.1098/rstb.2008.0309 http://rstb.royalsocietypublishing.org/content/364/1521/1301.full.pdf
68. Lorente De No, R. (1933) Vestibulo-ocular reflex arc. Arch. Neurol. Psychiat. (Chicago), 30:245-291.
69. Mandelbrot B (1967) How long is the coast of Britain? Statistical self-similarity and fractional dimension. Science 5(156):3775 pp. 636-638 DOI: 10.1126/science.156.3775.636 http://www.sciencemag.org/content/156/3775/636.abstract
70. Mandelbrot BB (1983) The fractal geometry of nature, Freeman, New York, ed. 2, 1983
71. Manfred R, Roy S (2001), (Quantum) space-time as a statistical geometry of fuzzy lumps and the connection with random metric spaces, Classical and Quantum.Grav. 18, p. 3039.
72. Mantegna RNSV, Buldyrev AL, Goldberger S, Havlin C, Peng K, Simons M, Stanley HE, (1994) Linguistic features of noncoding DNA sequences, Phys Rev Lett 73, pp. 3169–3172

73. Manto M (2008) The cerebellum, cerebellar disorders, and cerebellar research-two centuries of discoveries, Cerebellum. 7(4):505-16
74. Manto M, Marmolino D. (2009) Cerebellar ataxias. Curr. Opin. Neurol. 22(4):419-29 http://www.ncbi.nlm.nih.gov/pubmed/19421057
75. Marmolino D, Acquaviva F (2009) Friedreich's Ataxia: from the (GAA)n repeat mediated silencing to new promising molecules for therapy. The Cerebellum. 8(3):245-59
76. Marcer P.J. (1992) Order and chaos in DNA – the Denis Guichard Prizewinner: Jean-Claude Perez, in KIBERNETES 1992, vol 21, n°2, pp.60-61, ISSN 0368-492X http://www.emeraldinsight.com/journals.htm?articleid=1663901&show=abstract
77. Mardis ER (2006) Anticipating the $1000 genome. Genome Biol. 7(7):112. .doi: 10.1186/gb-2006-7-7-112
78. Marr D (1969) A theory of cerebellar cortex. J. Physiol., 202:437-470
79. Marr D (1982) Vision: A computational Investigation into the human representation and processing of visual information. New York: Freeman
80. Mattick JS [2001] Non-coding RNAs: the architects of eukaryotic complexity. EMBO reports 2(11):986–991 doi:10.1093/embo-reports/kve230 http://www.nature.com/embor/journal/v2/n11/full/embor291.html
81. Mattick JS (2004) The hidden genetic code of complex organisms. Scientific American, 291(4) pp. 60-67
82. Mattick JS (2005) The functional genomics of noncoding RNA. Science 309(5740) pp. 1527-1528 http://www.sciencemag.org/content/309/5740/1527
83. Mendel JG (1866). Versuche über PflanzenhybridenVerhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr, 1865 Abhandlungen:3–47. For the English translation, see: Druery CT and William Bateson (1901). "Experiments in plant hybridization". J of the Royal Horticultural Society 26: 1–32. Retrieved 2009-10-09
84. McCulloch WS, Pitts WH (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5:115-133
85. Meyerson M, Gabriel S, Getz G (2011). Advances in understanding cancer genomes through second-generation sequencing Nature Rev Gen 11, 685–696 | doi:10.1038/nrg2841
86. Minsky M, Papert S (1969) Perceptrons: An Introduction to computational geometry, The MIT Press, Cambridge MA, ISBN: 0 262 63022 2 http://www.amazon.com/Perceptrons-Marvin-Minsky/dp/B0000EG4U2/ref=sr_1_2?ie=UTF8&qid=1320482966&sr=8-2
87. Moruzzi G (1950) Problems in cerebellar physiology, Springfield, Ill. : C. C. Thomas, cop.
88. Neumann J (1958) The computer and the brain (Mrs. Hepsa Ely Silliman Memorial Lectures). Second Edition 2000, with Introduction by Patricia Churchland and Paul Churchland, Yale Univ. Press. http://www.barnesandnoble.com/w/the-computer-and-the-brain-john-von-neumann/1103017090
89. Oberdick J, Schilling K, Smeyne RJ, Corbin JG, Bocchiaro C, Morgan JI (1993) Control of segment-like patterns of gene expression in the mouse cerebellum. Neuron 10(6):1007-1018 doi:10.1016/0896-6273(93)90050-2 http://www.ncbi.nlm.nih.gov/pubmed/8318226
90. Ohno S (1972) So much ‘Junk DNA’’ in our genome. Brookhaven Symposia in Biology, (23):366-370 http://www.junkdna.com/ohno.html
91. Oller JW (2010) The antithesis of entropy: Biosemiotic communication from genetics to human language with special emphasis on the limmune systems. Entropy 12:631-705; doi:10.3390/e12040631 http://www.mdpi.com/1099-4300/12/4/631/pdf
92. Ozery-Flato M, Linhart C, Trakhtenbrot L, Izraeli S, Shamir R (2011) Large-scale analysis of chromosomal aberrations in cancer karyotypes reveals two distinct paths to aneuploidy Genome Biology 12:R61 doi:10.1186/gb-2011-12-6-r61
93. Pandolfo M (2008) Friedreich ataxia. Arch Neurol 65(10) http://archneur.ama-assn.org/cgi/reprint/65/10/1296.pdf
94. Pellionisz A, Szentágothai J (1973) Dynamic single unit simulation of a realistic cerebellar network model, 1973, Brain Res. , 49:83-99 Elsevier http://usa-siliconvalley.com/inst/pellionisz/ap_szj_72/ap_szj_72.html
95. Pellionisz AJ, Llinás R (1980) Tensorial approach to the geometry of brain function. Cerebellar coordination via a metric tensor. Neuroscience, 5:1761-1770. http://usa-siliconvalley.com/inst/pellionisz/80_metric/80_metric.html
96. Pellionisz AJ (1984) Coordination: A vector-matrix description of transformations of overcomplete CNS coordinates and a tensorial solution using the Moore-Penrose generalized inverse. J of Theoretical Biology, 110, 353-375. http://www.junkdna.com/pellionisz_coordination.pdf
97. Pellionisz A (1985) Tensorial aspects of the multidimensional approach to the vestibulo-oculomotor reflex and gaze. In: Reviews of Oculomotor Research. I. Adaptive Mechanisms in Gaze Control. (ed. by Berthoz A and Melvill-Jones G). Elsevier, Amsterdam, 281-296 http://usa-siliconvalley.com/inst/pellionisz/berthoz/berthoz.html
98. Pellionisz AJ (1985) Tensor network theory of the central nervous system and sensorimotor modeling. In Brain theory, Palm G, Aertsen A (Eds.) Berlin, Heidelberg, New York: Springer-Verlag 121-145.
99. Pellionisz AJ, Llinás R (1985) Tensor network theory of the metaorganization of functional geometries in the CNS. Neuroscience, 16, 245-273 http://www.junkdna.com/pellionisz_metaorganization_1985.pdf
100. Pellionisz A (1986) David Marr's theory of the cerebellar cortex: A model in brain theory for the "Galilean Combination of Simplification, Unification and Mathematization". In: Brain Theory (eds. Palm G., & Aertsen, A.), Springer Verlag, Berlin-Heidelberg-New York, 253-257 http://www.junkdna.com/pellionisz_marr.pdf
101. Pellionisz A (1987) Tensor network theory of the Central Nervous System. Encyclopaedia of Neuroscience (ed. G. Adelman), Birkhauser, p. 1196-1198 http://usa-siliconvalley.com/inst/pellionisz/encyclopaedia/encyclopaedia.html
102. Pellionisz A, Graf W (1987) Tensor Network Model of the "Three-Neuron Vestibulo-Ocular Reflex-Arc" in Cat. J. Theoretical Neurobiology, (5)127-151. http://www.junkdna.com/pellionisz_graf_1987.pdf
103. Pellionisz AJ (1989) Neural geometry: Towards a fractal model of neurons. In: Models of brain function, ed. By Cotterill, RMJ, Cambridge Univ. Press; 453-464 http://www.junkdna.com/pellionisz_fractal_purkinje_neuron_1989.pdf
104. Pellionisz AJ, LeGoff B, Laczko J, Berthoz A (1991) Multidimensional geometry intrinsic to head movements around distributed centers of rotation: A Neurocomputer Paradigm. In:The head-neck sensory-motor system/ Oxford Univ. Press. Eds. A. Berthoz, W.Graf and P. Vidal 117-125 http://books.google.com/books?id=6gLEx5slpOQC&pg=PA158&dq=pellionisz+legoff+laczko#v=onepage&q&f=false
105. Pellionisz AJ, Jorgensen CC, Werbos PJ (1992) Cerebellar neurocontroller project for aerospace applications. IJCNN International Joint Conf. on Neural Networks IEEE Catalog Number: 92CH3114-6 ISBN: Softbound Edition 0-7803-0559-0 http://usa-siliconvalley.com/inst/pellionisz/ijcnn92/ijcnn92.html
106. Pellionisz A (2002) FractoGene: Utility to use self-similar repetitions in the language-like genetic information as fractal sets. US Patent Application (Aug. 1st, 2002). In: Simons MJ, Pellionisz AJ (2006a), see also SF-gate, Nov.22; http://www.junkdna.com/plotkin.htm
107. Pellionisz A (2003) FractoGene: Design-tool for protein-based self-assembling nanostructures, materials and application. Invited Keynote Lecture at Nano-Bio Technology Session. Proc. Of the 204th Meeting of The Electrochemical Society, p. 1195
108. Pellionisz A (2006) PostGenetics: The journey of discovering “junk DNA”; Genetics beyond genes. Invited Keynote Lecture of European Inaugural Satellite Symposium of International PostGenetics Society, Proceedings of International Congress of Immunogenomics and Immunomics
109. Pellionisz A (2008a) The principle of recursive genome function. The Cerebellum, 7(3) 348-359, DOI 10.1007/s12311-008-0035-y http://ww.junkdna.com/pellionisz_principle_of_recursive_genome_function.pdf
110. Pellionisz A (2008b) Is IT ready for the dreaded DNA data deluge? http://www.youtube.com/watch?v=WJMFuc75V_w
111. Pellionisz A (2009a) From the principle of recursive genome function to interpretation of HoloGenome Regulation by Personal Genome Computers. Cold Spring Harbor Laboratory; Personal Genomes, Sept. 14-17 http://www.junkdna.com/pellionisz_csh.html
112. Pellionisz A (2009b) Personal genome computing: Breakthroughs, risks and opportunities, Churchill Club Panel http://www.youtube.com/watch?v=znowfg7WJ1Q
113. Pellionisz A (2010) Shop for your life - HolGenTech at PMWC2010 http://www.youtube.com/watch?v=mSRMCDCVg6Y
114. Perez JC (1988a) De nouvelles voies vers l'Intelligence Artificielle: pluri-disciplinarité, auto-organisation et réseaux neuronaux, Masson Paris, ISBN 2-225-81815-0.
115. Perez JC et al. (1988b) Fractal chaos: a new neural network holographic model - INNS conference Boston, USA. Neural Networks
116. Perez JC (1990a) Digital holograms computers, concepts and applications, in Neural networks: biological computers or electronic brains, Les entretiens de Lyon, Springer Verlag, ISBN 2-287-00051-8.
117. Perez JC (1990b) Integers neural network systems (INNS) using resonance properties of a Fibonacci’s chaotic golden neuron. Neural Networks 1, 859–865. IEEE 90CH2879-5. INSPEC Accession Number: 3926657 http://ieeexplore.ieee.org/Xplore/login.jsp?url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel2%2F148%2F3745%2F00137678.pdf%3Farnumber%3D137678&authDecision=-203
118. Perez JC (1991) Chaos, DNA, and neuro-computers: A golden link: The hidden language of genes, global language and order in the human genome. Speculations in Science and Technology 14, 336–346
119. Perez JC (1997) L’ADN décrypté, Résurgence, Liège Belgium. ISBN 2-87211-017-8

120. Perez JC (2008) Scale invariance embedded votes and self-emerging binary logics in the whole human genome. Relating the paper: What is complexity? by Gell-Mann PM. Complexity. 1(1) 1995, John Wiley and Sons http://www.scribd.com/doc/68476137/JcperezLiveBitsFromDNA-from-John-NASH-Logics-to-J-C-PEREZ-Fractals-Biomathematics
121. Perez JC (2009a) Codex Biogenesis. Resurgence, Liege Belgium. ISBN 2-87434-044-8. https://sites.google.com/site/codexbiogenesis/ and http://www.ei.ur.edu.uy/JCP_codex.pdf
122. Perez JC (2009b) Golden ratio and numbers in DNA http://golden-ratio-in-dna.blogspot.com and http://golden-ratio-in-dna.blogspot.com/2008/01/1997-perezs-generic-predictive-equation.html
123. Perez JC, Bertille JM (1990c) A spatio temporal novelty detector using Fractal Chaos model, IJCNN Conference Washington; Neural Networks, INNS USA
124. Perez JC (2010) Codon population in single-stranded whole human genome DNA are fractal and fine-tuned by the Golden Ration 1.618. Interdisciplinary Sciences: Computational Life Sciences Vol 2, Number 3, 228-340 DOI: 10.1007/s12539-010-0022-0 http://urbanshakedowns.files.wordpress.com/2011/01/adn-perez.pdf
125. Perez JC (2011a) Decoding non-coding DNA codes: Human genome meta-chromosomes architecture (support from Pr Luc Montagnier FMPRS World AIDS Foundation UNESCO and Jean-rené Fourtou Vivendi Universal chairman), BIT Life Sciences' Third Annual World Vaccine Congress-2011, Beijing March 2011 http://www.scribd.com/doc/57828784/jcperezBeijing032011
126. Perez JC (2011b) Caminos Interdisciplinaios, Seminario CLAVE_INTER, Espacio Interdisciplinario, Universidad de la Republica Montevideo Uruguay, 27 de Octubre 2011. http://www.ei.ur.edu.uy/enclave.html#confperez
127. Perez JC (2012) paper in preparation: DNA, Waveforms and Numbers: Unifying all 1496 HIV1 whole genomes
128. Peterson BW, Baker JA, Pellionisz AJ (1987) Multidimensional analysis of vestibulo-ocular and vestibulo-colllic reflexes (VOR and VCR). In: Proceedings of the International Symposium on Basic and Applied Aspects of Vestibular Function, Hong Kong
129. Peterson BW, Pellionisz AJ, Baker JA, Keshner EA (1989) Functional morphology and neural control of neck muscles in mammals. Am. Zoology, 29:139-149
130. Petoukhov S, He M (2010) Symmetrical analysis techniques for genetic systems and bioinformatics: Advanced patterns and applications. Medical Information Science Reference. ISBN 978-1605661247 http://www.flipkart.com/books/1605661244
131. Rigoutsos I (2006) BM Watson Research Center http://cbcsrv.watson.ibm.com/Tspd.html
132. Rigoutsos I, Floratos A, (1998) A. Combinatorial pattern discovery in biological sequences: The TEIRESIAS algorithm. Bioinformatics; 14(1) 55-67. pmid:9520502
133. Rigoutsos I, Huynh T, Miranda K, Tsirigos A, McHardy A, Platt D (2006) Short blocks from the noncoding parts of the human genome have instances within nearly all known genes and relate to biological processes. PNAS April 25, 2006 vol. 103 no. 17 6605-6610 http://www.pnas.org/content/103/17/6605.full
134. Roy S, Llinás R (2007) Dynamic geometry, brain function modeling, and consciousness. Progress in Brain Res. . Vol. 168, pp 133-144. doi:10.1016/S0079-6123(07)68011-X http://www.sciencedirect.com/science/article/pii/S007961230768011X
135. Schadt EE, Linderman MD, Sorenson J, Lee L, Nolan GP, Computational solutions to large-scale data management and analysis. (2010) Nature Rev. Genetics. (11) 647-657 http://www.nature.com/nrg/journal/v11/n9/full/nrg2857.html
136. Schrődinger E. (1944) What is life? Dublin Institute for Advanced Studies at Trinity College, Dublin, in February 1943 http://whatislife.stanford.edu/LoCo_files/What-is-Life.pdf
137. Shannon CE (1948) A Mathematical theory of communication, Bell System Technical Journal, 27, pp. 379–423 & 623–656
138. Shapshak P, Chiappelli, F, Commins D, Singer E, Levine AJ, Somboonwit C, Minagar A, Pellionisz, A. (2008) Molecular epigenetics, chromatin, and NeuroAIDS/HIV: Translational implications. Bioinformation. 3(1): 53–57. PMCID: PMC2586134 http://www.usa-siliconvalley.com/shapshak.pdf
139. Shepelyansky DL (2008) Fractal Weyl law for quantum fractal eigenstates. Physical Rev E 77, 015202 http://www.quantware.ups-tlse.fr/dima/myrefs/my165.pdf
140. Simons M, Pellionisz A (2006a) Genomics, morphogenesis and biophysics: Triangulation of Purkinje cell development. The Cerebellum. 5(1):27-35 http://www.junkdna.com/fractogene/05_simons_pellionisz.pdf
141. Simons M, Pellionisz A (2006b) Implications of fractal organization of DNA on disease risk genomic mapping and immune function analysis. Australasian and Southeast Asian Tissue Typing Association; 30th scientific meeting 22-24 November 2006, Chiangmai, Thailand http://www.junkdna.com/fractogene/06_simons_pellionisz.html
142. Stagnaro S (2011) Glycocalix quantum-biophysical-semeiotic evaluation plays a central role in demonstration of water memory-information http://www.sisbq.org/uploads/5/6/8/7/5687930/wmi_glycocalyx.pdf
143. Stagnaro S, Caramel S (2011) A new way of therapy based on water memory-information: the Quantum biophysical approach http://www.sisbq.org/uploads/5/6/8/7/5687930/qbtherapy.pdf
144. Sylvester JJ (1853) On a theory of the syzygetic relations of two rational integral functions, Comprising an application to the theory of Sturm's functions, and that of the greatest algebraical common measure, Philosophical Trans. of the Royal Society of London (The Royal Society) 143: 407–548, doi:10.1098/rstl.1853.0018, see also http://en.wikipedia.org/wiki/Covariant_vector and http://en.wikipedia.org/wiki/Covariance_and_contravariance_(computer_science)
145. Szentágothai J (1949) The elementary vestibulo-ocular reflex arc. J Neurophysiol. 13:(6)395-407 http://jn.physiology.org/content/13/6/395.long
146. Timchenko LT, Caskey CT (1999) Triplet repeat disorders: discussion of molecular mechanisms. Cell Mol Life Sci. 55(11):1432-47. http://www.ncbi.nlm.nih.gov/pubmed/10518991
147. Zipf GK (1949) Human behaviour and the principle of least-effort, Addison-Wesley, Cambridge MA
148. Zhuangzi (around the 4th century BC) Stanford encyclopedia of philosophy http://plato.stanford.edu/entries/taoism/
149. Venter C (2010) Multiple personal genomes await Nature 464, 676-677 (1 April 2010) | doi:10.1038/464676a; Published online 31 March 2010 http://www.nature.com/nature/journal/v464/n7289/full/464676a.html
150. Wang L, Brown SJ (2006) Bind N: A web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. Nucleic Acids Res., 34:W243-W248 and http://en.wikipedia.org/wiki/RNA_interference
151. Watson JD (2008) Editorial: The failed war on cancer, People against cancer. http://www.peopleagainstcancer.com/pdfs/news/20080916n2.pdf
152. Watson JD (2009) To fight cancer, know the enemy. New York Times August 6, A29 http://www.nytimes.com/2009/08/06/opinion/06watson.html?pagewanted=all
153. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG (et al) (2001) Science 291(5507):1304-1351. DOI: 10.1126/science.1058040, http://www.sciencemag.org/content/291/5507/1304.full
154. Weyl H (1912) Das asymptotisce lerteihingsgesetz der Eigenwerte linearer partieller Differentialgleichungen. Math. Ann. 71:441-479, see also in http://media.wiley.com/product_data/excerpt/04/35274083/3527408304.pdf
155. Wiener N (1948) Cybernetics or control and communication in the cnimal and the machine, Hermann & Cie Editeurs, Paris, The Technology Press, Cambridge, Mass., John Wiley & Sons Inc., New York; http://www.amazon.com/Cybernetics-Second-Control-Communication-Machine/dp/026273009X/ref=sr_1_1?ie=UTF8&qid=1319564993&sr=8-1
156. Yamashita S, Tsujino Y, Moriguchi K, Tatematsu M, Ushijima T (2006) Chemical genomic screening for methylation-silenced genes in gastric cancer cell lines using 5-aza-2'-deoxycytidine treatment and oligonucleotide microarray. Cancer Sci 97:64-71, http://onlinelibrary.wiley.com/doi/10.1111/j.1349-7006.2006.00136.x/pdf