Guenther Witzany *Editor*

Epigenetics in Biological Communication

Springer

Epigenetics in Biological Communication

Guenther Witzany
Editor

Epigenetics in Biological Communication

Springer

Editor
Guenther Witzany
Telos-Philosophische Praxis
Buermoos, Salzburg, Austria

ISBN 978-3-031-59285-0 ISBN 978-3-031-59286-7 (eBook)
https://doi.org/10.1007/978-3-031-59286-7

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Chapter 1
Epigenetics Integrates Development, Signaling, Context, RNA-Networks and Evolution

Guenther Witzany

Abstract The metamorphosis from larvae to adult butterflies has represented the "mystery" of life since the ancient Greeks. How could we explain the various steps of development from caterpillars to the most beautiful butterflies? A mystery prevalent in the twentieth century concerned the storage of the complete genetic information of an organism in the DNA of its every cell. How and why do so many different cell types develop throughout the lives of organisms at the right time and place? With the rise of epigenetics, the "fog" of mystery is starting to clear. We now know that the genetic storage medium in every living cell acquires an incredible plasticity through certain markings on the genetic text, linking the inheritable information with the contextuality of the real lifeworld of each organism. This means that the concrete living organism influences during development the various stages of gene expression, transcription, translation, immunity, and DNA repair actively and leads to various phenotypic outcomes without altering the DNA storage medium. More surprisingly, such variations of developmental phenotypes were found capable of successfully adapting to changing environmental circumstances. Better adapted organisms may lead to reprogramming in the epigenetic states which may reach an inheritable status for many generations or even become a fixed part of the genetic identity of the species. This is how evolution learns.

Introduction

Every cell, tissue, organ, and organism is competent to use signals for exchanging information, reaching common coordinations and organizations of both single cell and group behavior. Such signal-mediated interactions constitute biological communication (biocommunication).

G. Witzany (✉)
Telos-Philosophische Praxis, Buermoos, Salzburg, Austria
e-mail: witzany@sbg.at

G. Witzany (ed.), *Epigenetics in Biological Communication*,
https://doi.org/10.1007/978-3-031-59286-7_1

The regulatory system in charge of development, morphology, cell fate and identity, physiology, genetic instructions, immunity, memory/learning, and physical and mental health depends on epigenetic marks. The communication of cells, persistent viruses and their defectives such as mobile genetic elements, and RNA networks ensure both the transfer and reprogramming of regulatory instructions. But how are the different states of the epigenome orchestrated? This book will give an answer.

Epigenetic pathways respond to various signaling cues such as DNA methylation, histone variants, histone modifications, chromatin structure, nucleosome remodeling, and epigenetic interactions. Epigenetic signals are responsible for the establishment, maintenance, and reversal of transcriptional states that are fundamental to the cell's ability to memorize past events, such as changes in the external environment, sociosphere, or developmental cues. External signals trigger changes in the epigenome, allowing cells to respond dynamically, while internal signals direct activities necessary for body maintenance and tissue and organ repair.

With the emergence of epigenetic memory, organisms can fix historical and contextual impressive experiences (Shapiro 2014). Evolution from now on learned to learn. Such learning implies organisms can avoid always reproducing the same elements. This is key to adaptation.

However, inheritance of acquired characteristics is only one of the many examples of the explanatory power of epigenetics. Behavioral epigenetics demonstrates how environmental and social experiences produce individual differences in behavior, cognition, personality, and mental health and disease (Moore 2016, 2017; Chen et al. 2016; Punzi et al. 2018; Clayton et al. 2020; Skinner and Nilsson 2021).

Biological Communication as Main Characteristics of Life

Amid the twentieth century, cell–cell communication was recognized empirically, but it remained a side effect of cells functioning like machines, determined by the laws of physics and chemistry. Molecular biology specifically focused on the mechanistic details of cellular life. This changed dramatically when it was discovered that every interaction within or between cells depends on signaling pathways that are released and received. The goal of such signaling processes is a kind of coordinated behavior and/or the organization of something to reach a goal. In contrast to interactions of matter on abiotic planets where no signals are generated, sent, or received, signaling represents the most important feature in our planetary biosphere (Witzany 2015, 2019). Signals used in signaling processes may be molecules, electric impulses, tactile signals, or auditory and visual signals in higher eukaryotes. Signaling pathways are involved in any function within living cells such as gene expression, transcription, translation, immunity, and DNA repair in a variety of steps and substeps. Signaling pathways are essential for communication outside of the cell body, i.e., between cells, tissues, organs, and organisms. This has been proven throughout all domains of life (Witzany 2000, 2010, 2011, 2012a, 2013, 2017; Witzany and Baluška 2012a; Witzany and Nowacki 2016). Coordinated

behavior of at least two cells or more (like in tissues or organs) without signaling is impossible. If signaling functions well, the goals of communication can be reached, whereas if signaling is deformed or damaged, coordinated behavior of multiple living cells is disturbed or deformed, possibly resulting in disease or even the death of cells, tissue, organs, or organisms (Witzany 2020).

Even at the subcellular level, i.e., between viruses and RNA networks, signaling remains a major factor of coordinated behavior. In contrast to the cellular level viruses and RNA networks, communication depends on DNA sequence (genetic) identities that themselves represent the signal to be identified by a foreign interacting counterpart (Díaz-Muñoz et al. 2017; Sanjuán 2021). In the communication between RNA stem loops specifically, self- or non-self-identification is essential in determining whether binding occurs with a foreign sequence, leading to an increased RNA formation such as in ribonucleoprotein complexes (Stoddard and Belfort 2010; Sarkies and Miska 2013; Witzany 2014; Higgs and Lehman 2015). This means that at the quasispecies and RNA network level, no signaling molecules are generated, released, and received by the counterpart, but the nucleic acid sequence itself is the object of identification to determine whether it is relevant for interaction or not. Any definition of life must integrate both levels of communication: one in which signaling molecules are produced and one in which interacting partners themselves serve as signals for coordinating behavior (Villarreal and Witzany 2019).

Interestingly it has been found that whereas single RNA stem loops react according to physical chemical laws exclusively, if multiple RNA stem loops meet biological selection starts (Vaidya et al. 2012; Gwiazda et al. 2012; Manrubia 2022). A further result of research demonstrated that cooperative RNA stem-loops outcompete selfish ones (Stich et al. 2007; Villarreal and Witzany 2021, 2023).

All Organisms Communicate

Organisms actively compete for environmental resources. They assess their surroundings, estimate how much energy they need for particular goals, and then realize the optimum variant. They take measures to control certain environmental resources. They perceive themselves and can distinguish between “self” and “non-self.” Current empirical data on all domains of life indicate that unicellular organisms such as bacteria, archaea, giant viruses and protozoa as well as multicellular organisms such as animals, fungi and plants coordinate and organize their essential life functions through signaling processes. Signaling allows for real life coordination and organization and is a communicative action in which species-specific behavioral patterns and sign repertoires are used. Cells, tissues, organs and organisms that communicate share several key levels that are essential to all life forms and which serve as a uniform tool for investigating biological communication (Witzany 2016).

Keylevels of Biological Communication in the Cellular World

Research on biological communication identified several different levels of signal-mediated interactions:

(a) Sensing of abiotic circumstances such as light, gravity, temperature, water, dryness, wind, etc. Such environmental circumstances represent important indices on nutrition availability, symbiotic organisms, growth control, and developmental time clocks. They are not only sensed but monitored, and organisms of all domains store information about these indices in memory, to adapt better to repeated life situations.
(b) transorganismic communication is termed sign-mediated interaction between non-related species, as is the case in most symbiotic partnerships or in prey-predator interaction motifs. Such transorganismic communication processes can become very complex, as in the rhizosphere of plants (together with fungi, insects, bacteria, nematodes) and also in the normal ecosphere, such as the human mouth with its 500-plus different bacterial communities.
(c) Interorganismic communication is termed sign-mediated interaction processes between members of the same species or related species. They share species-specific vocabulary which may vary according to different ecospheres through dialects in which the signaling semantics differs according to adaptational needs.
(d) Intraorganismic communication: every organism consists of various parts that must be coordinated appropriately to install life functions. The signaling processes between these parts is termed intraorganismic communication. This also includes communication between DNA storage media and genetic parasites whether they are persistent settlers of viral origin or are defectives of infection events identified as mobile genetic elements and related RNA-networks, which now act as natural genetic engineers for better adaptation by the host organism (Witzany 2009; Fig. 1.1).

Context Determines Meaning

Whereas communication in general was explained by mathematical theories of language in the twentieth century, such explanatory models meanwhile are recognized as insufficient in explaining what really happened (Witzany 2015). According to such theories, if the grammar (syntax) of a signaling sequence is known, the meaning of the information-bearing molecule in the sequence can be identified easily (Shannon and Weaver 1949; Searls 2002; Nowak and Krakauer 1999). This was a misconception, because we now know that the context of use of the signaling sequence is crucial for the generation and even interpretation of meaning (Mead

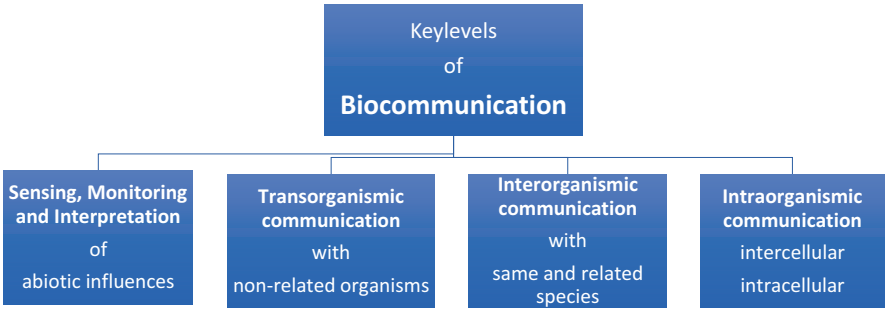

Fig. 1.1 Keylevels of biological communication in the cellular world

1934; Habermas 1994; Tomasello 2008). For example, if we take the phrase "the shooting of the hunters," it can be easily recognized that the syntax of this phrase does not determine the meaning which should lead to an appropriate response behavior by the receivers. The phrase may convey the meaning that the hunters are shot or the hunters are shooting, both of which will lead to different reactions. This was observed early by Ludwig Wittgenstein _ "The meaning of a word is its use" _ and Charles Sanders Peirce _ To identify the meaning we "have simply to look what habits it produces." (Wittgenstein 1953; Peirce 1923; Fig. 1.2).

The common coordination and organization of reactions depend on the commonly shared life world which ensures a sociological frame and the same interpretational background wherein signaling occurs.

When Karl von Frisch, who received noble prize 1973 for detection of the bee language, mixed two population of the same species of honey bees, one of upper Austria and one of Italy he observed that these two mixed bee populations struggled and fought each other.

Nutrition searching honey bees normally come back of their searching flights to the swarm and are dancing with various moving patterns to give the information to the swarm about (i) the direction in which the nutrition source can be found, (ii) the distance of this source and (iii) quality of found source. Experimentally-mixed colonies of Austrian and Italian bees revealed clear differences in the interpretation of the dance tempo, which indicates the distance to the feeding site. When the Austrian bees communicated a suitable feeding site at a distance of 300 m, for example, the Italian bees executed the instruction in exactly the right direction, yet over a distance of 500 m. Vice versa, a 200-m dance by the Italian bees meant a much shorter distance to the Austrian bees. Thus, despite identical rules being applied to the same linguistic signals, distinct differences existed in the meaning of the signs (von Frisch 1992). Interestingly, these differences in bee language dialects are even compatible over longer time distances. It depends on the capability of social learning of the bee populations. Longer time enables processes of training of different meanings of identical moving patterns (Su et al. 2008).

Fig. 1.2 Context determines meaning (e.g. "The shooting of the hunters") not syntax. Similarly the superficial grammar of DNA does not determine its meaning. The in vivo context which results in epigenetic markings represents a variable deep grammar which determines post-transcriptional modifications such as RNA editing and alternative splicing. Therefore in contrast to the opinions of Manfred Eigen, Sidney Brenner or Craig Venter algorithm-based DNA processing cannot generate both, superficial and deep grammar. (Witzany and Baluška 2012b; with Permission)

Natural Languages and Natural Communication Are Social Events

In linguistics, it has been known for long that linguistic and communication competence derive from everyday social interactions and social learning on how to correctly use words in utterances. This understanding dates back to the late Wittgenstein who noticed that rule-following is a social event, because "one cannot follow a rule only once."(Wittgenstein 1953).

Semiotician Charles Morris noticed that any kind of natural language is inherently interwoven with three kinds of behavioral rules (Morris 1946). Syntax denotes the combinatorial rules to correctly connect single signals to signal sequences such as characters to words and sentences. Pragmatics denotes the rules on how to use language within the correct context. Semantics denotes the rules on how to correctly combine signs with the designated object. If one level of rule is not followed, one cannot speak seriously in a natural language. Rule-following is learned through everyday social interactions, i.e., linguistic competence to correctly use a word is

combined with communicative competence, i.e., the ability to establish a social interaction (Austin 1975; Habermas 1994). This was the rationale behind placing language and communication within social sciences and distancing them from mathematical theories of language, because signal-mediated interactions within living populations are social events which cannot be assessed by physics, chemistry, or any other algorithm-based procedures.

Information theoretical sender-receiver narratives confuse their language of description with reality or how to interconnect thoughts, language and observations. This is the real problem with all metaphysical approaches such as objectivism, realism, and even ontology: not being aware, that the basic needs they describe do not depict physical objects (McCarthy 1984). I have outlined in previous works the reasons why none of the above-mentioned concepts can explain the evolution and function of languages used in communicative interactions (Witzany 1995, 2000, 2010). The core reasons are that these concepts cannot coherently explain:

- simultaneous understanding of identical meanings in two interacting partners, as expressed in successfully coordinated activity;
- differentiation between deep and superficial grammar of a statement along with differentiation between locutionary, illocutionary and perlocutionary speech acts;
- the de novo generation of coherent and context-dependent sentences.

Memory and Learning

Organisms that share the capability of storing information about experiences in the past have an actively generated background resource on which they can compare and evaluate more recent experiences in order to quickly or even better react than in previous situations. This is an essential competence for all reaction and adaptation purposes of living organisms. Such memory/learning skills can be found from akaryotes up to unicellular eukaryotes, fungi, animals and plants, although until recently, it had been mentioned only as a capability of higher animals. With the rise of epigenetics, the context-dependent marking of experiences at both the phenotype and the genotype level is an essential perspective to understand memory and learning in all organisms. Both memory and learning depend on a variety of successful communication processes within the whole organism.

Currently known epigenetic modifications depend on histone modifications – such as acetylation and deacetylation, methylation and demethylation, deamination, phosphorylation and dephosphorylation, isomerization, O-palmitoylation, ubiquitination and ADP-ribosylation – that determine the gene-expression processes. This represents a rich source of tools to mark experienced events of the organism on the genomic level (Atlasi et al. 2019). Epigenetic markings of certain chromosome sections to target memory relevant modes are essential for different identities of molecule groups, which represent the memorized identity as a kind of “frozen picture” of the total sum of biological communication processes of an organism in an

epigenetically relevant situational context. This means that the epigenetic marking of, for example, extraordinary stress situations – which activate all body parts and their dynamic interactional motifs represented in cells, tissues and organs – takes the "informational content" as the given relevant evaluation for imprinting processes (de Magalhães-Barbosa et al. 2022; Wu et al. 2024). But to evaluate or interpret memory, certain molecular identity groups must play relevant roles within the organism. This means they must trigger a different communication to the interconnected cellular tissues than the previous state where certain memory markings did not exist. If we look at the currently known facts on how organisms store experiences as a memory tool to learn how to better react and quickly adapt within the best energy-saving strategies, we can investigate an abundance of chemicals that serve as signaling molecules for coordination and organization of behavioral patterns. This means that not only memory and learning but all coordination and organization processes in organisms are the result of communicative interactions between cells, tissues and organs.

If an organism in real-life world context with its unique evolutionary and developmental history and identity is able to mark certain genetic setups that represent an environmentally determined specific replication pattern or transcription process, then memory is the result (Lickliter and Moore 2023). Memory marks a certain experienced event or multiple similar events to enable this organism to faster and/or more appropriate reaction, if similar situations occur (Moore 2023). This capability for better reaction may be termed successful "learning" of the organism based on this stored background information. The organism must differentiate between situations of the same structure without memory and with memory and then be able to evaluate the memory against stored background information (Thellier et al. 2018). This evaluation process may be termed "interpretation", as stored information leads to "learning", i.e. changing behavioral motifs such as faster/more appropriate reaction to similar real-life experiences. Evaluation of past experiences and comparison with present ones may lead to variable sensing, monitoring, evaluating and making decisions with far-reaching and differentiated consequences. In the long run, biological selection processes will lead to populations who represent an optimized memory/learning/interpretation competence.

Epigenetics Dynamically Links Genes to Environment, Experiences and Context

Although the DNA of every organism is stored in its every cell, cells differentiate in various forms to generate tissues and organs. This means a coordinated interaction happens within and between cells and among various cell types. During developmental stages as well as adulthood, if one cell of a certain tissue is damaged or dies, an appropriate new cell must develop instead at the right time in exactly the same position.

There are strong indicators that the formation of long-term memory in neural networks (in animals) is derived from epigenetic tagging via co-optation. This suggests that long-term memory in animal brains is an adaptational event, and cognitive memories, in principle, represent the same kind of epigenetic pathways as that of cellular memory (Levenson and Sweatt 2005; Day and Sweatt 2011). Epigenetic memory is modifiable in principle. As shown in the previous bee language example, such long-term memories can be modified through social learning over longer time periods.

Epigenetic markings memorize past experiences, especially if they have a distressing impact, e.g., stress, perinatal stress, malnutrition (hunger), strong experiences of pain, child abuse, or the absence of a mother's care (Bartlett et al. 2017; Kocamaz et al. 2022; Hoffmann et al. 2023). All these may lead to epigenetic markings that may even be inherited and result in risks of mental or metabolic diseases in later generations (Szyf 2021; Nilsson et al. 2022; Juruena 2023). This means that concrete social lifeworld experiences may alter the epigenetic status of gene expression and also influence marking on the germline cells (Spadafora 2017; Zhang et al. 2019; Conine and Rando 2022).

This also has consequences for the models in evolutionary biology. If an organism can better adapt to its social lifeworld or even to changing environmental conditions via learning processes through its memories of similar past experiences and accordingly change behaviors (or adapt metabolism), this trait may be inherited via epigenetic markings to the next generations (Fitz-James and Cavalli 2022; Fallet et al. 2023; Verdikt et al. 2023). Those generations will now share what the previous generations have generated – a new behavior or a new metabolic trait.

In the twentieth century, the key narratives in evolutionary biology were variation (error replication) and selection. However, inherited memorized and learned capabilities that led to better adapted phenotypes do not fit into the variations by genetic error replication (Frías-Lasserre and Villagra 2017). This means that the key narratives of evolutionary biology are insufficiently complex in integrating them. Consequently, an integrative theory of evolution must not only integrate the role of epigenetics but also that of viruses – as most biological entities on this planet, as well as the role of RNA networks in development and evolution.

Crucial Roles of Viruses in Spreading RNA-Networks

Besides the flexible epigenetic markings that are not part of heritable information transfer, transgenerational immune memory (siRNA, RNAinterference, CRIPRs/Cas) indicates that genetic parasite invasions that are warded off by the immune system will modify and mark those invasive genetic identities to be transferred as memory content via heredity to the offspring.

Our understanding of the key players in evolution and of the development of all organisms in all domains of life has been aided by current knowledge about RNA stem-loop groups, their proposed interaction motifs in an early RNA world and their

regulative roles in all steps and substeps of nearly all cellular processes, such as replication, transcription, translation, repair, immunity and epigenetic marking (Gesteland et al. 2006; Manrubia and Briones 2007; Ariza-Mateos et al. 2019). Cooperative evolution was enabled by promiscuous interactions between single-stranded regions in the loops of naturally forming stem-loop structures in RNAs (Briones et al. 2009; Shirogane et al. 2016). It was also shown that cooperative RNA stem-loops outcompete selfish ones and provide foundational self-constructive groups such as ribosome, editosome, spliceosome (Gwiazda et al. 2012).

From the beginning of nucleic acid sequence-based entities on this planet, the behavioral motif of genetic parasites is the driver of constant interactions – whether it be RNA viruses or similar RNA stem loop groups that are in constant interactions with other invading genetic parasites that must be identified, integrated as cooperative parts or warded off (Vaidya et al. 2012). Additionally, this interaction profile means identity problems to the RNA group, because it changes the genetic identity of the RNA group as well as that of the invaded agent (Villarreal 2009, Villarreal and Witzany 2015).

This may be disastrous if the former identity was successfully fixed and now may become irrelevant for the host organism, because the function cannot be continued. The new sequence order has to be identified as invasive species and as the relevant target to be warded off (Lambowitz and Zimmerly 2011). On the other side, this flexibility in identity features may cause the rise of a new and unexpected invasive agent identity, being a successful invader of formerly immune hosts (Villarreal 2012). This feature hints at a core feature of life and biotic planets: the constant and continued capability of RNA groups to resist or integrate novel genetic parasites, which drives (i) immune systems, (ii) genetic identities of host organisms and additionally (iii) genetic parasite identities in parallel.

For a long time, viruses have been considered as molecular invaders unable to replicate themselves. Meanwhile, it is more and more accepted that viruses have an abundance of genes not found in any cellular organism and are therefore older than cellular life. Several researchers have found that viruses, subviral networks and virus-derived parts (such as non-coding RNAs and mobile genetic elements) that are co-opted for host cellular needs play major roles in evolution and development of host organisms (Hayden and Lehman 2006; Smit et al. 2006; Zinad et al. 2017)). Prominent examples of coopted viral genes are the syncytin gene for placentation as well as the arc protein for neural plasticity (Villarreal 2009, 2016; Hantak et al. 2021).

Interestingly, short (miRNAs and siRNAs) and long non-coding RNAs and their derivatives, which can function as epigenetic marks of transcriptional gene silencing, also serve as defence tools against transposable elements and viruses (McKeown and Spillane 2014, Huang et al. 2014). Some researchers are of the opinion that the whole genetic content order of cellular organisms is determined by and regulated through such viral and subviral (defective) competencies (Villarreal 2005, 2015). This is because all viruses mark their genomes for self/non-self differentiation, e.g. the virus-first hypothesis suggests that epigenetic markings are transferred to cell-based organisms as infection-derived key competence of viruses that lead to innate and adaptive immune systems in all domains of life, which have been exapted and coopted for host purposes (Villarreal and Witzany 2010, 2013, 2018, 2019, 2023).

The Current RNAsphere

Molecular biology in the twentieth century assumed RNAs as intermediate molecules between DNA and proteins. Later on it was thought that non-coding RNAs are "junk", useless remnants of former evolutionary stages. This changed dramatically. RNAs are now being recognized as essential in all steps and substeps of gene regulation (Volff 2006; Regmi et al. 2022). RNAs are transferred into the life-world of cellular organisms via infection event (Villarreal and Witzany 2018).

Viruses represent the most abundant biological entities on this planet which outnumber living cells 10 times. All cells of all organisms of all domains of life are constantly interwoven into infection events from the beginning of their life until they are dying (Witzany 2012b). With every infection event the cellular organism is affected by RNA networks, that are sometimes exapted and co-opted for cellular needs (Koonin and Dolja 2013). RNA stem-loop groups represent an unmanageable quantity of sophisticated regulatory networks (Clark et al. 2013; Mattick and Amaral 2023; Ariza-Mateos et al. 2023). These groups are crucial in the following functions:

- DNA replication with important functions of centromeres and telomeres in genome maintenance
- RNA guidance of chromosome structure
- Regulation of transcriptional and post-transcriptional modifications by spliceosomes and editosomes according to the requirements of the context
- Regulatory pathways and coordination in all steps of the translation into proteins
- Epigenetic marking and short-term and long-term memory formation and its (re) modification
- DNA repair organization and coordination in all detailed steps and substeps
- Immunity organization and coordination in all steps and substeps by genome plasticity, V (ariable) D (iversity) J (oining) plasticity in adaptive immune response
- Genetic identity of organisms which initiates motifs of self-interaction or non-self-interaction
- Genetic content composition of host organisms by genetic parasites (viruses and defectives such as transposons and retroposons)
- Intron/exon genome fragmentation as a benefit in immune functions (CRISPR/Cas) as well as in genome modularity and complexity.

The active roles played by RNA stem-loop groups ensure all life processes currently known and start with the transcription process out of the relatively stable DNA storage medium (Villarreal and Witzany 2021). After transcription, an abundance of RNA stem-loop variants are available and interact in well-coordinated actions (Villarreal 2015). The folding loop remains as a binding prone single-stranded RNA sequence. Various motifs have been identified, yet all of them share a common function: they stabilize RNA tertiary formation. Such motifs include:

- pseudoknots, kissing loops, A-minor motifs, A-platforms, kink-turns, S-turns, tetraloops and their receptors, and a variety of non-canonical base-pairs and base-triples
- ribosomal frameshift as a natural technique to process alternative translation of an mRNA sequence by changing the open-reading frame
- bypassing translation
- competing endogenous RNAs.

All these highly coordinated and interconnected motifs of RNA stem-loop groups may alter the meaning of the information stored in the DNA according to the environmental and/or circumstantial requirements of an organism, which means that the information is context dependent (Nelson and Breaker 2017). This is documented by the essential roles of riboswitches also (Kavita and Breaker 2023).

In this respect RNA networks really represent the epicenter of genetic information. Epigenetic regulation of all steps and substeps of development are outlined by RNA mediated processes (Nowacki et al. 2011; Mattick 2009, 2023). Such programming of the development can even be modified in various ways triggered my experiences of the organism, such as stress situations, etc. as mentioned above.

Conclusions

The key narratives of the twentieth century in molecular biology, genetics, and evolutionary biology have changed dramatically. Several key assumptions have been proven wrong: (a) the central dogma of molecular biology (DNA-RNA-protein-anything else), (b) the consideration of non-coding DNA as "junk," (c) the one-gene-one-protein hypothesis, and (d) the understanding of viruses as mostly disease-causing genetic parasites. The mechanistic paradigm that genes function like Legos that can be inserted or deleted through genetic engineering without further consequences ignores the relevant roles of non-coding RNA regulatory networks and the epigenetic status of nucleotide sequences. With the knowledge that the context determines the meaning of signal sequences and the superficial sequence structure does not represent the hidden deep grammar of epigenetic markings, modifications, and programming, it became clear that the gene word order in organisms is not the result of random assemblies of nucleotides. In contrast, the genetic identities of organisms are mainly determined by the persistent invasions of genetic parasites, most of which remain as defectives (mobile genetic elements and other RNA-networks with repetitive sequence structure) that are exapted and co-opted for cellular needs. This shows us that the nucleotide sequences and genetic identities of organisms are largely the result of communication networks between viral clouds, RNA-networks, and cells, i.e., they are the result of social events. In conclusion epigenetics demonstrates how evolution learns through genes which communicate with the environment within the context of real life world of developing organisms.

References

Ariza-Mateos A, Briones C, Perales C, Domingo E, Gómez J (2019) The archaeology of coding RNA. Ann N Y Acad Sci 1447(1):119–134

Ariza-Mateos A, Briones C, Perales C, Sobrino F, Domingo E, Gómez J (2023, 2023) Natural languages and RNA virus evolution. J Physiol. https://doi.org/10.1113/JP284415

Atlasi Y, Megchelenbrink W, Peng T, Habibi E, Joshi O, Wang SY, Wang C, Logie C, Poser I, Marks H, Stunnenberg HG (2019) Epigenetic modulation of a hardwired 3D chromatin landscape in two naive states of pluripotency. Nat Cell Biol 21:568–578

Austin JL (1975) How to do things with words. Harvard University Press, London

Bartlett AA, Singh R, Hunter RG (2017) Anxiety and epigenetics. Adv Exp Med Biol 978:145–166

Briones C, Stich M, Manrubia SC (2009) The dawn of the RNA world: toward functional complexity through ligation of random RNA oligomers. RNA 15:743–749

Chen Q, Yan M, Cao Z, Li X, Zhang Y, Shi J, Feng GH, Peng H, Zhang X, Zhang Y, Qian J, Duan E, Zhai Q, Zhou Q (2016) Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. Science 351:397–400

Clark MB, Choudhary A, Smith MA, Taft RJ, Mattick JS (2013) The dark matter rises: the expanding world of regulatory RNAs. Essays Biochem 54:1–16

Clayton DF, Anreiter I, Aristizabal M, Frankland PW, Binder EB, Citri A (2020) The role of the genome in experience-dependent plasticity: extending the analogy of the genomic action potential. Proc Natl Acad Sci USA 117(38):23252–23260

Conine CC, Rando OJ (2022) Soma-to-germline RNA communication. Nat Rev Genet 23:73–88

Day JJ, Sweatt JD (2011) Cognitive neuroepigenetics: a role for epigenetic mechanisms in learning and memory. Neurobiol Learn Mem 96:2–12

de Magalhães-Barbosa MC, Prata-Barbosa A, da Cunha AJLA (2022) Toxic stress, epigenetics and child development. J Pediatr 98(Suppl 1):S13–S18

Díaz-Muñoz SL, Sanjuán R, West S (2017) Sociovirology: conflict, cooperation, and communication among viruses. Cell Host Microbe 22:437–441

Fallet M, Blanc M, Di Criscio M, Antczak P, Engwall M, Guerrero Bosagna C, Rüegg J, Keiter SH (2023) Present and future challenges for the investigation of transgenerational epigenetic inheritance. Environ Int 172:107776

Fitz-James MH, Cavalli G (2022) Molecular mechanisms of transgenerational epigenetic inheritance. Nat Rev Genet 23:325–341

Frías-Lasserre D, Villagra CA (2017) The importance of ncRNAs as epigenetic mechanisms in phenotypic variation and organic evolution. Front Microbiol 8:2483

Gesteland RF, Cech TR, Atkins JF (2006) The RNA world, 3rd edn. Cold Spring Harbor Laboratory Press, New York

Gwiazda S, Salomon K, Appel B, Müller S (2012) RNA self ligation: from oligonucleotides to full length ribozymes. Biochimie 94:1457–1463

Habermas J (1994) Actions, speech acts, linguistically mediated interactions and the lifeworld. Phil Probl Tod 1:45–74

Hantak MP, Einstein J, Kearns RB, Shepherd JD (2021) Intercellular communication in the nervous system goes viral. Trends Neurosci 44:248–259

Hayden EJ, Lehman N (2006) Self-assembly of a group I intron from inactive oligonucleotide fragments. Chem Biol 13:909–918

Higgs PG, Lehman N (2015) The RNA world: molecular cooperation at the origins of life. Nat Rev Genet 16:7–17

Hoffmann LB, McVicar EA, Harris RV, Collar-Fernández C, Clark MB, Hannan AJ, Pang TY (2023) Increased paternal corticosterone exposure influences offspring behaviour and expression of urinary pheromones. BMC Biol 21:186

Huang B, Jiang C, Zhang R (2014) Epigenetics: the language of the cell? Epigenomics 6:73–88

Juruena MF (2023) Early life stress, depression and epigenetics. Vitam Horm 122:307–337

Kavita K, Breaker RR (2023) Discovering riboswitches: the past and the future. Trends Biochem Sci 48:119–141
Kocamaz D, Franzke C, Gröger N, Braun K, Bock J (2022) Early life stress-induced epigenetic programming of hippocampal NPY-Y2 receptor gene expression changes in response to adult stress. Front Cell Neurosci 16:936979
Koonin EV, Dolja VV (2013) A virocentric perspective on the evolution of life. Curr Opin Virol 3:546–557
Lambowitz AM, Zimmerly S (2011) Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb Perspect Biol 3(8):a003616
Levenson JM, Sweatt JD (2005) Epigenetic mechanisms in memory formation. Nat Rev Neurosci 6:108–118
Lickliter R, Moore DS (2023) Molecular and systemic epigenetic inheritance: integrating development, genetics, and evolution. Hum Dev 1-13:305–317. https://doi.org/10.1159/000533192
Manrubia S (2022) The simple emergence of complex molecular function. Philos Trans A Math Phys Eng Sci 380(2227):20200422
Manrubia SC, Briones C (2007) Modular evolution and increase of functional complexity in replicating RNA molecules. RNA 13(1):97–107
Mattick JS (2009) Deconstructing the dogma. A new view of the evolution and genetic programing of complex organisms. Ann N Y Acad Sci 1178:29–46
Mattick JS (2023) A Kuhnian revolution in molecular biology: most genes in complex organisms express regulatory RNAs. BioEssays 45(9):e2300080
Mattick J, Amaral P (2023) RNA, the epicenter of genetic information: a new understanding of molecular biology. CRC Press, Abington
McCarthy T (1984) Translator's introduction. In: The theory of communicative action 1, ix. Beacon Press, Boston
McKeown PC, Spillane C (2014) Landscaping plant epigenetics. Methods Mol Biol 1112:1–24
Mead GH (1934) Mind, self, and society: from the standpoint of a social behaviorist. University of Chicago Press, Chicago, p 178
Moore DS (2016) Behavioral epigenetics. Wiley Interdiscip Rev Syst Biol Med 9(1)
Moore DS (2017) The developing genome: an introduction to behavioral epigenetics. Oxford University Press, Oxford
Moore DS (2023) On the evolution of epigenetics via exaptation: a developmental systems perspective. Ann NY Acad Sci 1529:21–32
Morris C (1946) Signs, language, and behavior. Prentice Hall, New York
Nelson JW, Breaker RR (2017) The lost language of the RNA world. Sci Signal 10(483):eaam8812
Nilsson EE, Ben Maamar M, Skinner MK (2022) Role of epigenetic transgenerational inheritance in generational toxicology. Environ Epigenet 8(1):dvac001
Nowacki M, Shetty K, Landweber LF (2011) RNA-mediated epigenetic programming of genome rearrangements. Annu Rev Genomics Hum Genet 12:367–389
Nowak MA, Krakauer DC (1999) The evolution of language. Proc Natl Acad Sci USA 96:8028–8033
Peirce CS (1923) How to make our ideas clear. In: Morris RC (ed) Chance, love, and logic: philosophical essays. Harcourt Brace & World, New York, pp 41–42
Punzi G, Bharadwaj R, Ursini G (2018) Neuroepigenetics of schizophrenia. Prog Mol Biol Transl Sci 158:195–226
Regmi R, Penton CR, Anderson J, Gupta VVSR (2022) Do small RNAs unlock the below ground microbiome-plant interaction mystery? Front Mol Biosci 9:1017392
Sanjuán R (2021) The social life of viruses. Annu Rev Virol 8:183–199
Sarkies P, Miska EA (2013) Molecular biology. Is there social RNA? Science 341:467–468
Searls DB (2002) The language of genes. Nature 420:211–217
Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana

Shapiro JA (2014) Epigenetic control of mobile DNA as an interface between experience and genome change. Front Genet 5:87
Shirogane Y, Watanabe S, Yanagi Y (2016) Cooperative interaction within RNA virus mutant spectra. Curr Top Microbiol Immunol 392:219–229
Skinner MK, Nilsson EE (2021) Role of environmentally induced epigenetic transgenerational inheritance in evolutionary biology: unified evolution theory. Environ Epigenet 7(1):dvab012
Smit S, Yarus M, Knight R (2006) Natural selection is not required to explain universal compositional patterns in rRNA secondary structure categories. RNA 12:1–14
Spadafora C (2017) Sperm-mediated transgenerational inheritance. Front Microbiol 8:2401
Stich M, Briones C, Manrubia SC (2007) Collective properties of evolving molecular quasispecies. BMC Evol Biol 7:110
Stoddard B, Belfort M (2010) Social networking between mobile introns and their host genes. Mol Microbiol 78:1–4
Su S, Cai F, Si A et al (2008) East learns from west: Asiatic honeybees can understand dance language of European honeybees. PLoS One 3(6):e2365
Szyf M (2021) Perinatal stress and epigenetics. Handb Clin Neurol 180:125–148
Thellier M, Luettge U, Norris V, Ripoll C (2018) Plant accommodation to their environment: the role of specific forms of memory. In: Baluska F et al (eds) Memory and learning in plants. Springer, Cham
Tomasello M (2008) Origins of human communication. MIT Press, Cambridge, MA
Vaidya N, Manapat ML, Chen IA, Xulvi-Brunet R, Hayden EJ, Lehman N (2012) Spontaneous network formation among cooperative RNA replicators. Nature 49:72–77
Verdikt R, Armstrong AA, Allard P (2023) Transgenerational inheritance and its modulation by environmental cues. Curr Top Dev Biol 152:31–76
Villarreal LP (2005) Viruses and the evolution of life. ASM Press, Washington
Villarreal LP (2009) The source of self: genetic parasites and the origin of adaptive immunity. Ann N Y Acad Sci 1178:194–232
Villarreal L (2012) Viruses and host evolution: virus-mediated self identity. Adv Exp Med Biol 738:185–217
Villarreal LP (2015) Force for ancient and recent life: viral and stem-loop RNA consortia promote life. Ann N Y Acad Sci 1341:25–34
Villarreal LP (2016) Viruses and the placenta: the essential virus first view. APMIS 124:20–30
Villarreal LP, Witzany G (2010) Viruses are essential agents within the roots and stem of the tree of life. J Theor Biol 262:698–710
Villarreal LP, Witzany G (2013) The DNA habitat and its RNA inhabitants: at the dawn of RNA sociology. Genom Ins 6:1–12
Villarreal LP, Witzany G (2015) When competing viruses unify: evolution, conservation, and plasticity of genetic identities. J Mol Evol 80:305–318
Villarreal LP, Witzany G (2018) Editorial: genome invading RNA-networks. Front Microbiol 9:581
Villarreal LP, Witzany G (2019) That is life: communicating RNA networks from viruses and cells in continuous interaction. Ann NY Acad Sci 1447:5–20
Villarreal LP, Witzany G (2021) Social networking of quasi-species consortia drive virolution via persistence. AIMS Microbiol 7:138–162
Villarreal L, Witzany G (2023) Self-empowerment of life through RNA networks, cells and viruses. F1000Res 12:138
Volff JN (2006) Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. BioEssays 28:913–922
von Frisch K (1992) Decoding the language of the bee. In: Jan L (ed) Nobel lectures, physiology or medicine 1971–1980. World Scientific Publishing, Singapore, pp 76–87
Wittgenstein L (1953) Philosophical investigations. Basil Blackwell, Oxford
Witzany G (1995) From the, logic of the molecular syntax to molecular pragmatism. Explanatory deficits in Manfred Eigen's concept of language and communication. Evol Cogn 1:148–168.

Witzany G (2000) Life: the communicative structure. Books on Demand, Norderstedt
Witzany G (2009) Non-coding RNAs: persistent viral agents as modular tools for cellular needs. Ann NY Acad Sci 1178:244–267
Witzany G (2010) Biocommunication and natural genome editing. Springer, Dordrecht
Witzany G (ed) (2011) Biocommunication in soil microorganisms. Springer, Heidelberg
Witzany G (ed) (2012a) Biocommunication of fungi. Springer, Dordrecht
Witzany G (ed) (2012b) Viruses: essential agents of life. Springer, Dordrecht
Witzany G (ed) (2013) Biocommunication of animals. Springer, Dordrecht
Witzany G (2014) Language and communication as universal requirements for life. In: Kolb V (ed) Astrobiology: an evolutionary approach. CrC Press, Boka Raton
Witzany G (2015) Life is physics and chemistry and communication. Ann NY Acad Sci 1341:1–9
Witzany G (2016) Key levels of biocommunication. In: Gordon R, Seckbach J (eds) Biocommunication: sign-mediated interactions between cells and organisms. World Scientific, Singapore, pp 37–61
Witzany G (ed) (2017) Biocommunication of archaea. Springer, Dordrecht
Witzany G (2019) Communication as the main characteristic of life. In: Kolb V (ed) Handbook of astrobiology. CrC Press, Boka Raton, pp 91–105
Witzany G (2020) What is life? Front Astron Space Sci 7:7
Witzany G, Baluška F (2012a) Biocommunication of plants. Springer, Berlin/Heidelberg
Witzany G, Baluška F (2012b) Life's code script does not code itself. The machine metaphor for living organisms is outdated. EMBO Rep 13:1054–1056
Witzany G, Nowacki M (eds) (2016) Biocommunication of ciliates. Springer, Dordrecht
Wu Z, Qu J, Zhang W, Liu GH (2024) Stress, epigenetics, and aging: unraveling the intricate crosstalk. Mol Cell 84:34–54
Zhang Y, Shi J, Rassoulzadegan M, Tuorto F, Chen Q (2019) Sperm RNA code programmes the metabolic health of offspring. Nat Rev Endocrinol 15:489–498
Zinad HS, Natasya I, Werner A (2017) Natural antisense transcripts at the interface between host genome and mobile genetic elements. Front Microbiol 8:2292

Chapter 2
Epigenetic Control of Cell Fate Decisions by Enhancer-Derived Long Noncoding RNAs

John S. Mattick

Abstract The development of mammals from conception to adult involves trillions of cell fate decisions that must be made with 4-dimensional precision to form the myriad of architecturally sculpted and intricately connected bones, muscles and organs. The spatiotemporal control of this developmental symphony resides in genetic loci termed enhancers, of which ~1 million exist in the human genome. Enhancers have long been thought to act as sites for the binding of transcription factors that are brought into contact with target gene promoters by chromatin looping. However, enhancers are, in fact, genes that produce alternatively spliced and often modified long noncoding RNAs. Enhancer-derived RNAs act as scaffolds to recruit chromatin modifying proteins and transcription factors containing intrinsically disordered regions (which are subject to a multitude of posttranslational modifications) into phase separated domains for the feed forward control of developmental ontogeny.

Introduction

It should come as no surprise that most of the human genome is devoted to the specification of development, although most molecular research has been focused on biochemistry, physiology and the genetic factors that cause illness: debilitating mutations in protein-coding sequences and more subtle regulatory variations associated with complex traits and diseases, which lie overwhelmingly in the vast intronic and 'intergenic' regions of the genome.

Since the 1960s, the understanding of the information held in the human genome has been viewed almost entirely through the lens of microbial genetics, established in the 1960s with the *lac* operon and the 'genetic code'. In bacteria and other simple

J. S. Mattick (✉)
School of Biotechnology and Biomolecular Sciences, UNSW Sydney, Sydney, NSW, Australia
e-mail: j.mattick@unsw.edu.au

G. Witzany (ed.), *Epigenetics in Biological Communication*,
https://doi.org/10.1007/978-3-031-59286-7_2

organisms subjected to genetic analysis, mainly of metabolic pathways, most genes were shown to encode proteins, whose expression is regulated by transcription factors that bind to cis-acting DNA sequences to promote or repress transcription of the 'messenger' RNA intermediate. The 'noncoding' genes specified ribosomal and transfer RNAs involved in translation. Small nucleolar RNAs and small nuclear RNAs discovered later in eukaryotes were also viewed as infrastructural, involved in chemical modification of rRNAs and tRNAs, and splicing, respectively, although there were also others of unknown function, such as vault RNAs, Y RNAs, and 7SK and 7SL RNAs (Mattick and Amaral 2022).

Throughout, the abiding but usually unstated assumptions were that genes are generally synonymous with proteins and that the mechanisms that control multicellular development are limited to those that regulate bacterial physiology, using 'transcription factors' and other proteins that recognize target sequences in DNA or RNA (Buchler et al. 2003; Howard and Davidson 2004; Peter and Davidson 2016). Genetic analyses of developmental mutations in multicellular organisms, historically the plant *Arabidopsis thaliana*, the fruitfly *Drosophila melanogaster*, mouse and human, and more recently the nematode *Caenorhabditis elegans*, were generally interpreted in the same framework, aided by a combination of phenotypic, technical and interpretative biases at a time when molecular data was largely unavailable (Mattick 2009). Even the discovery of small regulatory RNAs (microRNAs and related species in the RNA interference pathway) was viewed as an elaboration of translational control, rather than the portent of a wider system of RNA regulation (Mattick and Amaral 2022).

Theoretical biologists predicted that most of the human genome would not encode proteins on the basis that, if the human genome had a similar gene density and mutation frequency as bacteria, the 'mutational load' would be unbearable. The unjustified following conclusion was that most of the genome must be nonfunctional (Ohno 1972), despite the debates between the Mendelians and the quantitative trait geneticists, ignoring the possibility that regulatory sequences have different structure-function relationships and phenotypic spectra in developmentally complex organisms (Mattick and Amaral 2022; Pheasant and Mattick 2007).

It was little shock then that genome sequencing showed that only 2% of the human genome specified mRNAs, the rest again assumed to be mainly nonfunctional, relying on circular assessments of the 'neutral' evolution rate, which assumed that transposon-derived sequences common to mammalian genomes ('ancient repeats') are inert (Pheasant and Mattick 2007; Waterston et al. 2002; Christmas et al. 2023), despite abundant evidence to the contrary (Peaston et al. 2004; Lowe et al. 2007; Tsirigos and Rigoutsos 2009; Kelley et al. 2014; Davis et al. 2017; Trizzino et al. 2017; Sundaram and Wysocka 2020; Bartonicek et al. 2022).

The Anomalies

There were unexpected and troubling observations (Mattick 2023a): The number of protein-coding genes does not increase with developmental complexity, illustrated by the fact that *C. elegans*, with only ~1000 somatic cells (a few muscle cells, nerves and a gut) and humans with an estimated 30–40-trillion cells arranged onto a myriad of precisely sculpted and architecturally diverse bones, muscles and organs, and a brain with ~100 billion neurons, both have ~20,000 protein-coding genes (Hillier et al. 2005; Amaral et al. 2023), termed the g-value enigma (Hahn and Wray 2002). Moreover, many of the encoded proteins have similar functions. By contrast, most of the noncoding sequences are differentially transcribed to produce a plethora of intronic, antisense and intergenic noncoding RNAs at different cell types and developmental stages (Okazaki et al. 2002; Kapranov et al. 2002; Rinn et al. 2003; Bertone et al. 2004; Cheng et al. 2005; Kapranov et al. 2005; Carninci et al. 2005; Katayama et al. 2005; Furuno et al. 2006; Ravasi et al. 2006; Kapranov et al. 2007a, b), across complex interleaved loci with fuzzy boundaries (Kapranov et al. 2007b; Mattick 2003; Engstrom et al. 2006; Mattick and Makunin 2006; Gingeras 2007), the extent of which increases with developmental complexity (Taft et al. 2007; Liu et al. 2013). The later ENCODE studies showed that noncoding regions of the genome are replete with dynamic epigenetic marks and other biochemical signatures of function (Birney et al. 2007; Dunham et al. 2012).

The noncoding transcripts were, nonetheless, often dismissed as transcriptional noise or technical artefacts, supported by their low expression levels and evolutionary conservation relative to protein-coding sequences, but in reality reflecting high cell specificity and low stoichiometry of the transcripts, low sequencing depth (Clark et al. 2011; Mercer et al. 2012; Deveson et al. 2017; Wu et al. 2021a) and the rapid evolution of regulatory regions (Pheasant and Mattick 2007; Frith et al. 2006), including those identified as 'enhancers' (Kapusta et al. 2013; Johnson and Guigo 2014; Fueyo et al. 2022) (see below), the main source of adaptive radiation (Levine and Davidson 2005; Andolfatto 2005; Carroll 2008; Jeong et al. 2008; Hong et al. 2008; Feschotte 2008; Shubin et al. 2009; Villar et al. 2015; Cosby et al. 2021).

In parallel, the superficial but widely accepted response to the lack of increase in the numbers of genes encoding transcription factors and other regulatory proteins between nematodes and humans was to assert that the power of 'combinatorial' control is sufficient to enable "*a dramatic expansion in regulatory complexity*" (Smith and Valcarcel 2000; Levine and Tjian 2003). The assertion was not formalized theoretically, mathematically or mechanistically, and did not distinguish between multiple inputs at the point of regulatory decision and sequential action in a decisional hierarchy, although some later models considered both (Peter and Davidson 2011).

Meanwhile, intensive analyses of particular genomic regions, notably the *bithorax* complex in *Drosophila* and the haemoglobin gene cluster in mammals, revealed the existence of loci that control the pattern of expression of the nearby protein-coding genes, termed 'cis-regulatory domains' in the former (Bender 2020) and 'locus control regions' in the latter (Fraser and Grosveld 1998). Molecular analyses showed that the *bithorax* cis-regulatory domains comprise separate transcriptional units that express long non-coding RNAs (lncRNAs) (Bender 2020) and are associated with an unusual genetic phenomenon, termed 'transvection' (or 'allelic cross-talk') (Judd 1988), likely the signature of compound heterozygotes using trans-acting RNA signals (Micol and García-Bellido 1988; Micol et al. 1990; Mattick and Gagen 2001). Indeed, confocal imaging showed that regulatory domains in the *Drosophila Abd-B* locus located on one chromosome can associate with the *Abd-B* Hox (protein-coding) gene located on the other, the direct visualization of transvection (Ronshaugen and Levine 2004).

Similar analyses also exposed extensive 'intergenic' transcription in the human beta-globin locus (Ashe et al. 1997) (and others (Rogan et al. 2004; Jones and Flavell 2005; Rinn et al. 2007)), associated with a similar genetic phenomenon, termed 'transinduction' (Ashe et al. 1997). Arrays of highly conserved noncoding sequences were also discovered in the vicinity of key developmental genes (Bejerano et al. 2004; Sandelin et al. 2004), in some cases acting as enhancers (Woolfe et al. 2005; de la Calle-Mustienes et al. 2005; Pennacchio et al. 2006; Visel et al. 2008; Nolte et al. 2014; Dickel et al. 2018; Snetkova et al. 2022).

Enhancers

The term 'enhancer' was coined in 1981 in relation to a sequence that could, in an orientation-independent manner, increase the expression of a distal cloned beta globin gene (Banerji et al. 1981). The term was soon adopted to refer to genetic loci that direct the spatiotemporal patterns of gene expression during development, which include the examples above. Many enhancer loci were then identified by genetic and bioinformatic approaches: the former by insertions of reporter genes, called 'enhancer trapping' (O'Kane and Gehring 1987; McCall et al. 1994; Springer 2000; Trinh and Fraser 2013), and CRISPR/Cas9-mediated cellular screens (Gasperini et al. 2019); the latter by genome-wide analysis of the binding positions of presumed signature proteins (the 'transcriptional co-activators' P300 and Mediator) and correlated histone modifications (Heintzman et al. 2007, 2009; Creyghton et al. 2010; Shen et al. 2012; Pradeepa et al. 2016; Henriques et al. 2018; Chen and Liang 2020), nucleosome-depleted regions and/or the expression of 'enhancer RNAs' (eRNAs) (De Santa et al. 2010; Wang et al. 2011a; Wu et al. 2014; Andersson et al. 2014; Shlyueva et al. 2014; Kim et al. 2015a, b; Arner et al. 2015; Chen et al. 2017; Sartorelli and Lauberth 2020), which yield somewhat different prediction sets and blur the distinctions between enhancers and protein-coding genes (Henriques et al. 2018; De Santa et al. 2010; Shlyueva et al. 2014; Core et al.

2014; Heidari et al. 2014; Young et al. 2017; Rickels and Shilatifard 2018; Grossman et al. 2018; Tippens et al. 2018; Halfon 2019; Osmala and Lähdesmäki 2020).

Enhancers are associated with a range of developmental processes, including, among others, the emergence of multicellularity and phenotypic diversity (Rubinstein and de Souza 2013; Schwaiger et al. 2014; Sebé-Pedrós et al. 2016; Gaiti et al. 2017; Wong et al. 2020), *Drosophila* body segment specification (Perry et al. 2011; Smith and Shilatifard 2014), vertebrate skeletal, limb and craniofacial development (Nolte et al. 2014; Park et al. 2004; Monge et al. 2003; Prabhakar et al. 2008; White et al. 2021), neuronal development and differentiation (Miyagi et al. 2006; Sikorska et al. 2008; Closser et al. 2021; Mangan et al. 2022), the recent evolution of primates (Glinsky and Barakat 2019), the increase in human thumb size and rotation (Prabhakar et al. 2008), and developmental disorders (Smith and Shilatifard 2014; Long et al. 2020; Armendariz et al. 2023).

Multiple enhancers ensure precise patterns of gene expression during body plan specification (Perry et al. 2011; Woltering and Duboule 2010; Lagha et al. 2012), not just of mRNAs (including those encoding transcription factors (Setten et al. 2021)) but also lncRNAs potentially involved in feed-forward enhancer action (Yang et al. 2016; Soibam 2017) (see below). There are clusters of enhancers dubbed 'super-enhancers', 'stretch enhancers', 'enhancer jungles' or 'nested enhancers' (Chen and Liang 2020; Hnisz et al. 2013; Whyte et al. 2013; Parker et al. 2013; Pott and Lieb 2015; Wang et al. 2019; Li and Ovcharenko 2020; Thomas et al. 2021) and 'shadow enhancers' that may suppress noise, provide robustness and/or fine-scale regulation (Hong et al. 2008; Cannavò et al. 2016; Waymack et al. 2020; Kvon et al. 2021; Lin et al. 2022). Enhancers have been described as "*information integration hubs*" in development (Buecker and Wysocka 2012) and early evidence suggested that the entire enhancer sequence, not just its promoter (see below) is essential to fulfil its spectrum of functions (Kioussis and Festenstein 1997; Sipos et al. 1998). The number of enhancers that exist in the mammalian genome is estimated to be at least several hundred thousand and likely in excess of one million (Dunham et al. 2012; Shen et al. 2012; Chen and Liang 2020; De Santa et al. 2010; Andersson et al. 2014; Heidari et al. 2014; Thurman et al. 2012; Zhu et al. 2013; Wang et al. 2018a; Arnold et al. 2020).

Indeed, an enormous amount of information must be required to orchestrate human ontogeny from the point of conception to a functioning adult of approximately 30–40 trillion cells (Bianconi et al. 2013; Hatton et al. 2023), which involves a binary tree of ~10^{14} cell fate decisions (divide and/or differentiate to a lineage-specific or terminal state) that accurately specifies the design and connectivity of each bone, muscle and organ, and is so precise that monozygotic twins are phenocopies. This number of cell fate decisions is 4 orders of magnitude greater than the digital information content of the human genome, which begs the question of how the information is compressed and decompressed.

In 1986, in an attempt to reconcile enhancer action with transcription factor control of gene expression, it was proposed that enhancers comprise clusters of transcription factor binding sites that are brought into contact with target gene promoters by DNA looping (Ptashne 1988), a model that has since permeated textbooks and

the literature (Levine et al. 2014). The persistence of the initial interpretation of how enhancers work has been referred to a "*founder fallacy*" maintained by "*validation creep*" (Halfon 2019).

There is compelling evidence that enhancer action leads to chromatin reorganization and the juxtaposition of enhancers with target genes in topologically-associated domains (Ronshaugen and Levine 2004; Tolhuis et al. 2002; Deng et al. 2012; Rao et al. 2014; Bonev et al. 2017; Furlong and Levine 2018; Souaid et al. 2018; Popay and Dixon 2022; Golov et al. 2023). There is also abundant evidence that transcription factors bind the promoters of enhancers and other genes specifying lncRNAs (Dunham et al. 2012; Grossman et al. 2018; Whyte et al. 2013; Cawley et al. 2004; Mattioli et al. 2019).

However, there is no evidence that the transcription factors bound at enhancers are brought into direct contact with the promoters of 'target' genes (Popay and Dixon 2022; Benabdallah et al. 2019), nor any rationalization of the logic of a two-step process (i.e., why the transcription factors do not address their presumed target promoters in the first place).

Transcription of Enhancers

By contrast, there is substantial evidence that enhancers are transcribed in the cells in which they are active (Wang et al. 2011a; Andersson et al. 2014; Kim et al. 2015a; Arner et al. 2015; Li et al. 2016), not only producing short bidirectional RNAs (like protein-coding genes) but also multi-exonic lncRNAs (Mattick et al. 2023; Mattick 2023b). Indeed, enhancers exhibit all of the characteristics of genes that express an RNA product, albeit not translated into a protein. Consistent with their role in fine-scale developmental control, enhancer lncRNA expression is more tissue specific than that of protein-coding genes (Heidari et al. 2014; Mattioli et al. 2019; Cabili et al. 2011).

There is also increasing molecular and genetic evidence that the lncRNAs expressed from enhancers are integral to their action (Yang et al. 2016; Arnold et al. 2020; Wang et al. 2011b; Maass et al. 2012; Lai et al. 2013; Kretz et al. 2012, 2013; Li et al. 2013; Melo et al. 2013; Sauvageau et al. 2013; Lam et al. 2014; Xiang et al. 2014; Sun et al. 2014; Paralkar et al. 2014; Hacisuleyman et al. 2014; Pefanis et al. 2015; Cajigas et al. 2015; Goff et al. 2015; Pnueli et al. 2015; Yin et al. 2015; Lai et al. 2015; Aguilo et al. 2016; Deng et al. 2016; Isoda et al. 2017; Alvarez-Dominguez et al. 2017; Alexanian et al. 2017; Werner et al. 2017; Shii et al. 2017; Cajigas et al. 2018; Morrison et al. 2018; Groff et al. 2018; Tsai et al. 2018; Hua et al. 2018; Fatima et al. 2019; Tan et al. 2020; Carullo et al. 2020; Gao et al. 2020; Andergassen and Rinn 2021; Allou et al. 2021; Cipriano et al. 2021; Cajigas et al. 2021; Oh et al. 2021; Zibitt et al. 2021; Tan and Marques 2022; Chignon et al. 2022; Lewis et al. 2022; Shiau et al. 2022; Li et al. 2023), although there are contradictory reports (Paralkar Vikram et al. 2016; Engreitz et al. 2016). Genomic regions associated with complex traits and diseases are enriched in enhancers (Harrison and Bose 2022) and express a multitude of lncRNAs (Bartonicek et al. 2017; Hardwick et al. 2019).

Phase Separated Domains and Intrinsically Disordered Regions

The mechanisms of action of enhancer lncRNAs are vague, if not murky, but appear to involve the formation of phase-separated chromatin domains in conjunction with proteins containing intrinsically disordered regions (IDRs) (Arnold et al. 2020; Cabili et al. 2015; Lin et al. 2015; Hnisz et al. 2017; Sabari et al. 2018; Hahn 2018; Cho et al. 2018; Shrinivas et al. 2019; Sanchez de Groot et al. 2019; Morf et al. 2020; Ahn et al. 2021; Wu et al. 2021b; Somasundaram et al. 2022).

Transcription and splicing occur in phase separated domains (Cho et al. 2018; Tripathi et al. 2012; Boehning et al. 2018; Chong et al. 2018a; Lu et al. 2018; Quintero-Cadena et al. 2020; Shao et al. 2022), and many lncRNAs, including those identified as originating from enhancers, appear to localize to such domains (Wu et al. 2021b; Pessina et al. 2019). Proteins containing IDRs are present in and essential for the function of nearly all proteins involved in gene regulation during development, including RNA polymerase (Quintero-Cadena et al. 2020), most transcription factors (Minezaki et al. 2006; Liu et al. 2006; Guo et al. 2012; Staby et al. 2017; Boija et al. 2018; Wang et al. 2021), homeodomain proteins (Robertson et al. 2018; Basu et al. 2020), histones and histone modifying proteins (Peng et al. 2012; Lazar et al. 2016), subunits of PRC2 and other chromatin-modifying complexes (Plys et al. 2019; Lee et al. 2023), chromatin-binding proteins (Watson and Stott 2019; Musselman and Kutateladze 2021), 19 of the 26 subunits in the human Mediator complex (Tóth-Petróczy et al. 2008; Nagulapalli et al. 2016; Richter et al. 2022), RNA binding proteins (Beckmann et al. 2015; Xiao et al. 2019; Hentze et al. 2018), splicing factors (Korneta and Bujnicki 2012; Chen and Moore 2014), nuclear hormone receptors and cell signalling proteins (Tantos et al. 2012; Wright and Dyson 2015), and many proteins involved in cancer (Deiana et al. 2019) and neurological functions (Bakthavachalu et al. 2018; Maharana et al. 2018; White et al. 2019; Zhang et al. 2020; Loganathan et al. 2020; Yu et al. 2021).

The number of proteins containing IDRs, and the number of repeated motifs within them, correlate with developmental complexity (Beckmann et al. 2015; Niklas et al. 2015, 2018; Balcerak et al. 2019; Dunker et al. 2015; Yruela et al. 2017). The majority of proteins subject to alternative splicing contain IDRs (Niklas et al. 2018), and IDRs are overrepresented in alternatively spliced exons subject to tissue- and lineage-specific regulation (Romero et al. 2006; Buljan et al. 2012; Ellis et al. 2012; Barbosa-Morais et al. 2012; Gueroussov et al. 2017), which change the function and subcellular localization of the isoform and modulate phase transitions within the cell (Weatheritt et al. 2012; Weatheritt and Gibson 2012). IDRs have been shown to direct transcription factor binding specificity *in vivo (*Brodsky et al. 2020*)* and are major sites of post-translational modifications (Darling and Uversky 2018; Bah and Forman-Kay 2016).

IDRs, including those in transcription factors, bind RNA (Castello et al. 2016; Järvelin et al. 2016; Basu and Bahadur 2016; Oksuz et al. 2023), often through arginine-rich motifs punctuated by glycine residues that form G-quadruplex

structures, the frequency and order of which likely determines the affinity, flexibility and selectivity of the motifs for their target RNAs (Hentze et al. 2018; Ozdilek et al. 2017). Other motifs containing tyrosine residues flanked by glycine and/or serine residues have similar function in recognizing RNA sequences (Hentze et al. 2018).

Mutations that cause human monogenic diseases occur more commonly in IDRs than globular domains, indicating that, despite their superficially simple composition, IDRs have strong sequence constraints (Castello et al. 2013; Meyer et al. 2018). Expansions of repeat motifs in IDRs that alter phase separation are also associated with developmental and neurological disorders (Basu et al. 2020; White et al. 2019; Shin and Brangwynne 2017; Jain and Vale 2017; Hofmann et al. 2019; Elbaum-Garfinkle 2019; Tsang et al. 2020; Zbinden et al. 2020).

RNA Scaffolding of Phase Separated Domains

RNA is the structural scaffold of phase-separated domains (Shevtsov and Dundr 2011; Zhang et al. 2015; Fay and Anderson 2018; Polymenidou 2018; Garcia-Jove Navarro et al. 2019; Frank and Rippe 2020; Roden and Gladfelter 2021; Quinodoz et al. 2021; Quinodoz and Guttman 2022). It has been proposed that lncRNAs play a central role in organizing the 3-dimensional genome, including the formation of spatial compartments and transcriptional condensates, and hence the 4-dimensional patterns of gene expression during differentiation and development (Hnisz et al. 2017; Quinodoz et al. 2021; Quinodoz and Guttman 2022; Mao et al. 2011; Henninger et al. 2021; Luo et al. 2021; Mele and Rinn 2016).

Many lncRNAs bind 'promiscuously' to the repressive chromatin-modifying complex PRC2 (Davidovich et al. 2013; Davidovich and Cech 2015; Wang et al. 2017; Kang et al. 2020), reflecting the ability of PRC2 to interact with many partners determined by local concentration, alternative splicing and post-transcriptional and posttranslational modifications (see below). Reciprocally, IDRs have been described as having promiscuous, i.e., multilateral, interactions, which enable "*plasticity and pliability in RNA–protein complexes … essential for complex, precise and fine-tuned regulation*" (Balcerak et al. 2019; Cumberworth et al. 2013; Protter et al. 2018; Macossay-Castillo et al. 2019).

It has been shown that phase separation drives chromatin looping (Ahn et al. 2021) and is required for the action of enhancers and super-enhancers (Hnisz et al. 2017; Sabari et al. 2018; Hahn 2018; Shrinivas et al. 2019; Nair et al. 2019); that transcription factors activate genes by forming phase separated domains with RNA polymerase II (Chong et al. 2018a; Boija et al. 2018); that Mediator and RNA polymerase II associate in transcription-dependent condensates (Sabari et al. 2018; Cho et al. 2018; Chong et al. 2018a; Boija et al. 2018; Ramasamy et al. 2023) and that phase separated domains scaffolded by lncRNAs, including repeat-derived RNAs, mediate heterochromatin formation (Hofmann et al. 2019; Quinodoz et al. 2021; Strom et al. 2017; Falk et al. 2019; Rawal et al. 2019; Williams et al. 2020; Novo