



## Opinion

# Diagrams Describing the Evolution of Gene Expression, the Emergence of Novel Cell Types During Evolution, and *Evo-devo*



Andrei P. Kozlov<sup>1,2\*</sup>

<sup>1</sup>Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia; <sup>2</sup>Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Received: June 02, 2023 | Revised: June 25, 2023 | Accepted: August 01, 2023 | Published online: September 01, 2023

In 1958, Francis Crick<sup>1</sup> formulated the sequence hypothesis and central dogma of molecular biology.<sup>1</sup> The formulation of the sequence hypothesis was, as follows: “The specificity of a piece of nucleic acid is expressed solely by the sequence of its bases, and this sequence is a (simple) code for the amino acid sequence of a particular protein.” The formulation for the central dogma was, as follows: “This states that once ‘information’ has passed into protein, it cannot get out again. In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but the transfer from protein to protein, or from protein to nucleic acid is impossible. Information means here the precise determination of sequence, either of bases in nucleic acids or amino acids in the protein.”<sup>1</sup>

In 1970, F. Crick suggested a diagram that described the central dogma of molecular biology (Fig. 1).<sup>2</sup> According to F. Crick,<sup>2</sup> the arrows in the diagram represent “the directional flow of detailed, residue-by-residue, sequence information from one polymer molecule to another.” “Solid arrows show general transfers; dotted arrows show special transfers.” However, as stated by F. Crick, “it says nothing about what machinery of transfers is made of,” and “it says nothing about control mechanisms.” Since then, the deciphering of the machinery of information transfer and control mechanisms has become the agenda of molecular biology.

Some of the important results of this work included the theory of regulatory genome and gene regulatory networks,<sup>3</sup> the realization of the increase in regulatory complexity during the evolution of multicellular organisms,<sup>4,5</sup> and the understanding of the self-organizing

principles of genome organization and function.<sup>6</sup> However, the authors considered that the determination of how the genome structure correlates to function remains an open question.<sup>7</sup> Furthermore, the validity of the sequence hypothesis has recently been challenged based on the assessment of the peptide-to-protein folding process, and the public health concerns raised by the genomic approach to health and disease.<sup>8</sup> The author of the present study adds that we are still further away from understanding how new functions and morphological novelties originate during progressive evolution, and what the role of gene expression is in this process. The present study was devoted to this problem. It appears that the fundamentals of molecular biology need to be revised, and that new paradigms are needed to address the unresolved biological and medical problems.

The author of the present study is working on a new theory of *carcino-evo-devo*, or the theory of the evolutionary role of hereditary tumors.<sup>9</sup> According to the main hypothesis of the *carcino-evo-devo* theory, hereditary tumors contribute to the emergence of new cell types, tissues, and organs by providing extra cell masses for the expression of evolutionarily novel genes and gene combinations. A special chapter of the *carcino-evo-devo* theory was devoted to *carcino-evo-devo* diagrams that described the main postulates of the theory.

The three diagrams used in the present study, which were obtained from previous articles published by the author, are the multilevel reciprocity diagram obtained from a reference (Fig. 2),<sup>10</sup> and the *carcino-evo-devo* diagrams obtained from another reference (Fig. 3).<sup>11</sup> In Figure 2, the processes of evolution at different levels of structural organization were interconnected using feedback arrows. That is, these are mutually interdependent. The multilevel reciprocity diagram was coined to explain the neutralization of competitive interrelations between genes that emerged during genome evolution at the cellular and multicellular level. The diagrams that described the evolution of development (*evo-devo*) according to the *carcino-evo-devo* theory are presented in Figure 3.

According to the *carcino-evo-devo* theory, tumor-bearing organisms (Carcino) represent the transitional forms in the evolution of development. Normal ontogenies cannot directly participate in progressive evolution. This is prohibited by the *carcino-evo-devo* diagram (the lack of Devo→Evo arrows). The unfolded *carcino-evo-devo* diagram describes the four successive steps in the progressive evolution of development (marked by colored arrows). Kozlov<sup>11</sup> provided

**Keywords:** Central dogma; Evolution; Gene Expression; Diagrams.

**Abbreviations:** Carcino, ontogenies with neoplastic development; *carcino-evo-devo*, coevolution of hereditary tumors with normal development, the theory of the evolutionary role of hereditary tumors; CL, cellular level of structural organization; Devo, normal ontogenies; *evo-devo*, evolution of development; Evo, progressive evolution of ontogenies; MML, macromolecular level of structural organization; MCL, multicellular level of structural organization; *TSEEN* genes, tumor specifically expressed, evolutionarily novel genes.

\*Correspondence to: Andrei P. Kozlov, Vavilov Institute of General Genetics, Russian Academy of Sciences, No. 3, Gubkina Street, 117971 Moscow, Russia. ORCID: <https://orcid.org/0000-0003-4611-1534>. Tel: +7 (911) 173-2792, E-mail: [contact@biomed.spb.ru](mailto:contact@biomed.spb.ru)  
**How to cite this article:** Kozlov AP. Diagrams Describing the Evolution of Gene Expression, the Emergence of Novel Cell Types During Evolution, and *Evo-devo*. *Gene Expr* 2023;000(000):000–000. doi: 10.14218/GE.2023.00031.

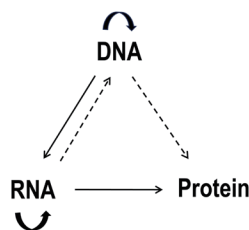


Fig. 1. The diagram describes the central dogma of molecular biology.<sup>2</sup>



Fig. 2. The multilevel reciprocity diagram.<sup>10</sup> CL, cellular level of structural organization; MCL, multicellular level of structural organization; MML, macromolecular level of structural organization.

a more detailed description of the *carcino-evo-devo* diagrams.

In the present study, the author suggested the modification of the diagram of the central dogma to describe the evolution of gene expression, drew a cellular diagram to describe the origin of new cell types, and used these together with Figures 2 and 3 to construct a formula that describes the evolution of gene expression, the origin of new cell types, *evo-devo*, and the complexity growth during progressive evolution.

The diagram construction was based on the ideology of central dogma and *carcino-evo-devo* theory, and on the analysis of empirical data and essential biological connections at and among the macromolecular, cellular and multicellular levels of structural organization.

### Modification of the central dogma diagram: the diagram that describes the evolution of gene expression

The striking similarity between Figures 1 and 3a suggested that the central dogma diagram can be unfolded upwards along the evolutionary scale, in a manner similar to that presented in Figure 3b. The author notes that the dotted arrow RNA→DNA in Figure 1 may describe the origin of evolutionarily new genes by retroposition. Thus, the unfolded central dogma diagram would indeed describe the genome, transcriptome and proteome evolution, and may be constructed in the following manner (Fig. 4).

Similar to Figure 3b, which describes the four successive steps in the evolution of development (*evo-devo*), Figure 4 describes the four successive steps (marked by colored arrows) in genome evolution, and the evolution of gene expression. Each successive step in the genome evolution in Figure 4 may involve a number of evolutionarily novel genes that participate in the origin of new differentiated cell types and morphological novelties.

### Construction of the formula for the coevolution of gene expression and *evo-devo*

Since Figures 3b and 4 describe the four successive steps in the evolution of gene expression and *evo-devo*, the author attempted to construct a single formula that described the coevolution of these processes. The processes described in Figure 4 belong to the macromolecular level of organization. The processes described in Figure 3b belong to the multicellular level of organization. Figures 3b and 4 can be connected to Figure 2 using the math symbol of subset  $\subset$  to obtain the formula for the coevolution of gene expression and *evo-devo* at two levels (MML and MCL) in the course of progressive evolution (Fig. 5).

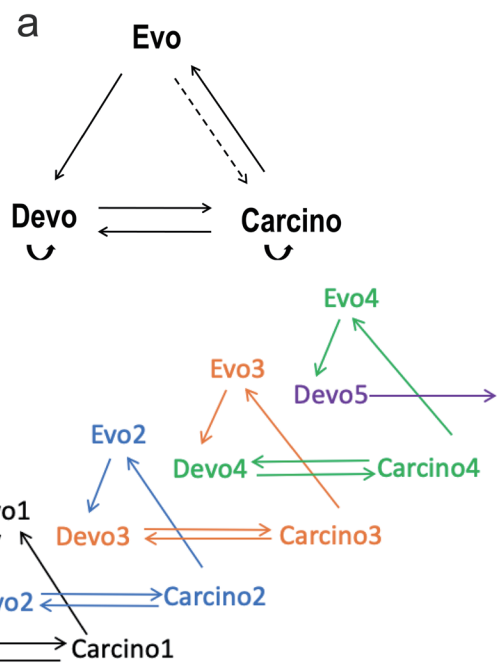
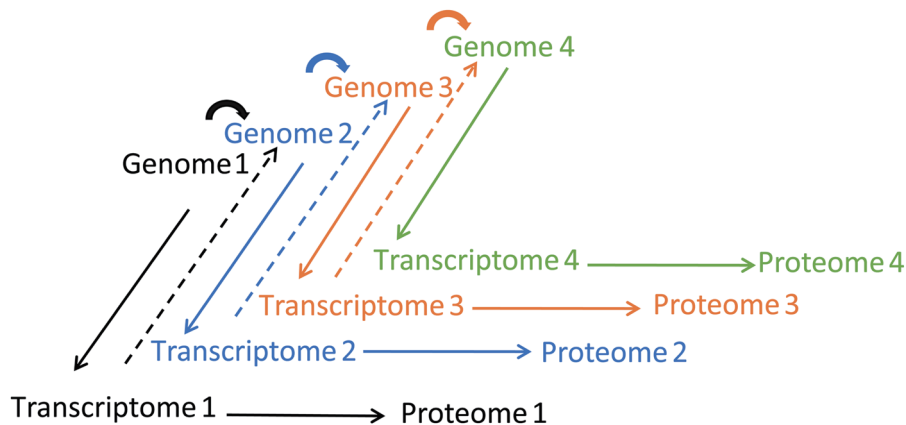


Fig. 3. *Carcino-evo-devo* diagrams.<sup>11</sup> (a) The regular *carcino-evo-devo* diagram. The arrows in the *carcino-evo-devo* diagram indicate the participation in the corresponding process, or essential connections. The evolutionary influence on tumor development (e.g., anti-cancer selection) is represented by the dotted arrow (evo→carcino). Curved arrows were added to the *carcino-evo-devo* diagram 3a, as compared to the corresponding diagram in a reference,<sup>11</sup> in order to designate the capability of self-replication that is necessary for participation in evolutionary processes. (b) The *carcino-evo-devo* diagram unfolded upwards along the evolutionary scale. This presents the four successive steps in progressive evolution (Evo 1, Evo 2, Evo 3, and Evo 4), with the participation of different hereditary tumors (Carcino 1, Carcino 2, Carcino 3, and Carcino 4), generating four new ontogenies (Devo 2, Devo 3, Devo 4, and Devo 5) with evolutionarily novel progressive traits. Merely arrows important for the evolution of development are shown. Carcino, ontogenies with neoplastic development; Devo, normal ontogenies; Evo, progressive evolution of ontogenies.

Thus, Figure 5 describes the four consecutive steps in the coevolution of gene expression and *evo-devo*. The evolution of development (*evo-devo*) resulted in evolutionary innovations and morphological novelties, which constitute the complexity growth.

### Cellular diagrams

Figure 5 suggests that one diagram is missing, that is, the cellular diagram. The author constructed similar triangle diagrams for the cellular level of organization (Fig. 6). Stem/embryonal cells are cells that participate in normal development, differentiated cells are terminally differentiated cells, and tumor cells include cancer stem cells and the hierarchy of parenchymal cells at different stages of (abnormal) differentiation. The arrows in Figure 6 represent the possible transitions among stem/embryonal, tumor and differentiated cells. The embryonal→differentiated arrow represents the main route of development towards terminally differentiated cells. The embryonal→tumor arrow designates the possibility of the tumor origin from embryonal cells.<sup>12</sup> The tumor→differentiated arrow represents the capability of tumor cells to differentiate with the loss of malignancy.<sup>13</sup> The tumor→embryonal arrow represents



**Fig. 4. The diagram describes the genome evolution and evolution of gene expression (transcriptome and proteome evolution).** The dotted transcriptome1→genome2 and similar arrows represent the origin of evolutionarily novel genes by retroposition (RNA-mediated mechanisms), and the curved genome 1 → genome 2, and similar arrows represent the origin of evolutionarily novel genes by DNA-mediated mechanisms. Merely arrows important for the evolution of gene expression are shown.

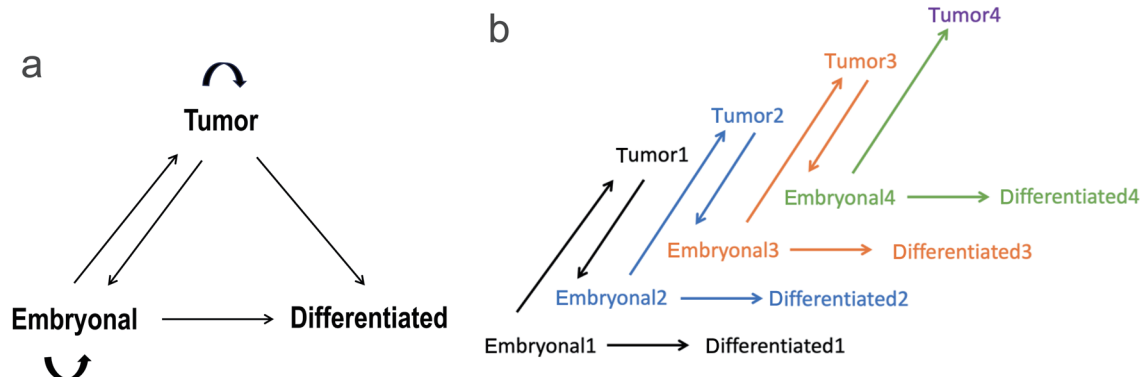


**Fig. 5. Formula for the coevolution of gene expression and evo-devo.** The  $\supset$  is a symbol that refers to a subset. This symbol was flipped according to the position on the diagram.

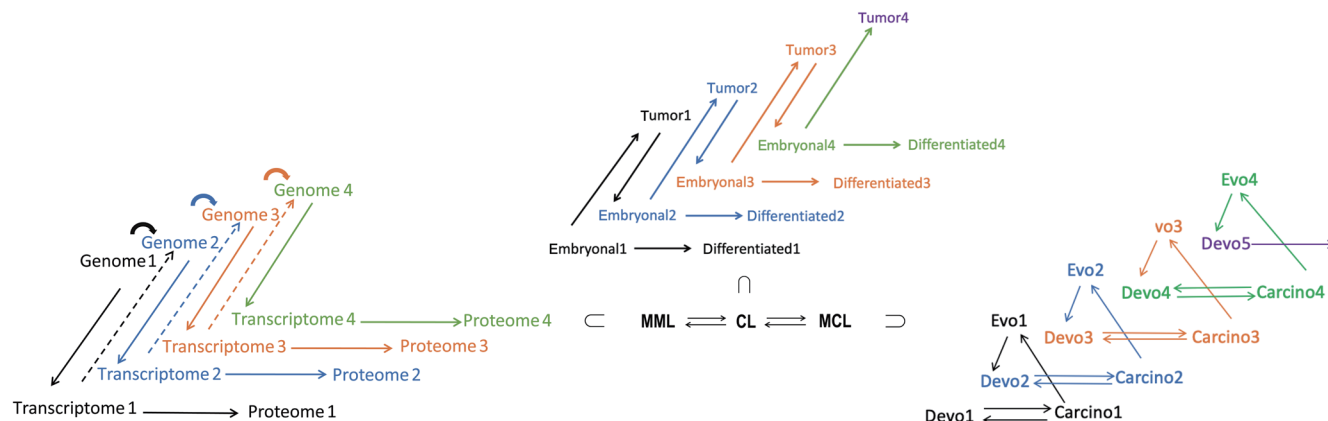
the main postulate of the *carcino-evo-devo* theory. That is, the possibility that hereditary tumor cells would become an integral part of normal development after the expression of evolutionarily novel genes and gene combinations, and the acquisition of new functions.<sup>11,13</sup> The curved arrows represent the capability of stem cells and cancer stem cells to replicate. The absence of arrows from terminally differentiated cells means that these have very little (or no)

capacity to replicate, generate tumors, or de-differentiate (although the author is aware of the reprogramming and trans-differentiation experiments, and that few examples exist in nature).<sup>14–17</sup> Thus, the omitted arrows represent the prohibited transitions.

In summary, Figure 6a describes the possible transitions among stem/embryonal, tumor and differentiated cells, and Figure 6b describes the origin of new differentiated cell types, and the incorpora-



**Fig. 6. Cellular diagrams.** (a) The regular cellular diagram describes the possible transitions among stem/embryonal, tumor, and differentiated cells; (b) The cellular diagram unfolded upwards along the evolutionary scale, describing the origin of new differentiated cell types during evolution. The arrows represent the possible transitions among stem/embryonal, tumor, and differentiated cells. Merely arrows important for the origin of new cell types in progressive evolution are shown. differentiated, terminally differentiated cells; embryonal, stem/embryonal cells; tumor, tumor cells.



**Fig. 7. The complete unfolded formula for the evolution of gene expression, the origin of new cell types, *evo-devo*, and the multilevel complexity growth in progressive evolution.** The  $\cap$  is the symbol for a subset. This symbol was flipped according to the position on the diagram. Merely arrows important for the evolution of gene expression, the origin of new cell types, *evo-devo*, and the complexity growth in progressive evolution are shown.

tion of hereditary tumor cells into normal development in the course of progressive evolution. The number of differentiated cell types in multicellular organisms is considered a measure of complexity.<sup>13</sup>

### The complete unfolded formula for the evolution of gene expression, the origin of new cell types, *evo-devo*, and the multilevel complexity growth in progressive evolution

Using Figure 5 and the unfolded Figure 6b, the complete unfolded formula for the evolution of gene expression, the origin of new differentiated cell types, *evo-devo*, and the multilevel complexity growth in progressive evolution was constructed (Fig. 7). Figure 7 describes the four consecutive steps in the evolution of gene expression, the origin of new differentiated cell types, *evo-devo*, and the multilevel complexity growth in progressive evolution. For simplicity, Figure 7 may be drawn in a regular folded form (Fig. 8).

### Evolution of gene expression as part of a multilevel evolutionary process

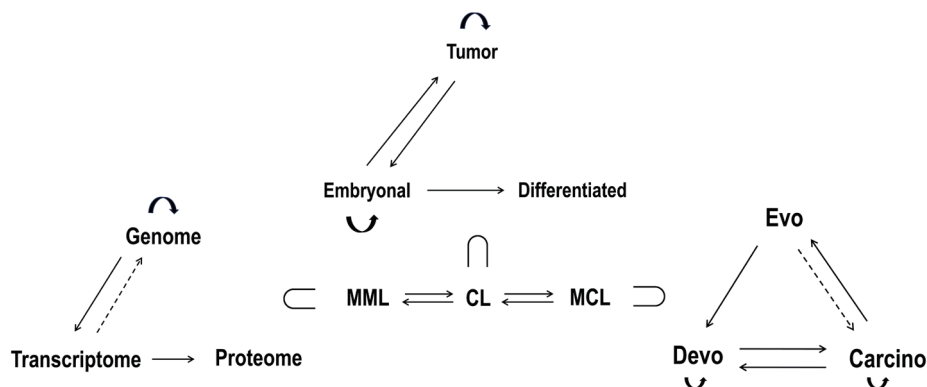
In previous publications, the author investigated the evolution of gene expression against the background of genome evolution, and the increase in gene number in genomes of evolving organisms.<sup>10,18</sup>

The advanced treatment of competitive interrelations between genes revealed that when the gene number increased in evolving genomes, the enforcement of gene competition and appearance of antagonistic relations between genes took place. The evolution process would result in the neutralization of incompatibility between genes through the disconnection of incompatible gene products at both the cellular and multicellular level.<sup>10,18</sup> The functional organization of the genome in groups of compatible genes, with antagonistic relations between at least some genes of different compatibility groups, would follow after such coevolution.<sup>10</sup> Each transcriptome or proteome in Figure 4 corresponds to the result of the expression of compatible genes. The various transcriptomes (transcriptome 1, transcriptome 2, etc.) or proteomes (proteome 1, proteome 2, etc.) in Figure 4 may include the products of incompatible genes.

Figure 2 presents the relationship between evolutionary processes at different levels, and the reciprocity between these. In the figure, the processes of evolution at different levels of structural organization were interconnected using feedback arrows, that is, these are mutually interdependent.<sup>10</sup>

### The role of tumors in the evolution of gene expression

In a series of publications that followed, the author formulated the hypothesis of evolution by tumor neofunctionalization.<sup>13</sup> Accord-



**Fig. 8. The complete regular folded formula for the evolution of gene expression, the origin of new cell types, *evo-devo*, and the multilevel complexity growth in progressive evolution.**



ing to this hypothesis, hereditary tumors provide additional cell masses for the expression of evolutionarily novel genes (which originate in the DNA of germ cells, but not in tumor cells) and gene combinations that may lead to the origin of new cell types, tissues, and organs in evolution.

This is the first time that the hypothesis of evolution by tumor neofunctionalization considered the genome evolution and evolution of gene expression in the interaction with processes of evolution at the cellular and multicellular level, and the role of tumors in this interaction. This hypothesis explains what previously formulated hypotheses (evolution by gene duplication<sup>19</sup> and the genetic theory of morphological evolution<sup>20</sup>) could not: the source of cells, in which evolutionarily novel genes or gene combinations are expressed, and thereby participate in the origin of new cell types, tissues and organs.

The role of tumors in the evolution of gene expression was confirmed through the discovery of a new class of tumor-specifically expressed, evolutionarily novel (*TSEEN*) genes in the laboratory of the author. *TSEEN* genes are expressed in a broad range of tumors, but are not expressed or weakly expressed in normal cells, including embryonic and stem cells.<sup>21,22</sup> The author proposed that *TSEEN* genes can be considered as a new superclass of novel and evolving genes that are only expressed in tumors. There are several classes and families of *TSEEN* genes that include *TSEEN* genes from different phyla of organisms.<sup>22</sup>

### Carcino-evo-devo theory

With the accumulation of supporting evidence and experimental data from the author's laboratory and other laboratories, the hypothesis of evolution by tumor neofunctionalization grew into the theory of the evolutionary role of hereditary tumors. The content of the theory was published in a series of theoretical articles by the author,<sup>9</sup> and as a monograph in English,<sup>13</sup> Russian,<sup>23</sup> and Chinese<sup>24</sup> languages.

The major postulates of the *carcino-evo-devo* theory are, as follows:<sup>9</sup> (1) tumors participate in the evolution of development; (2) hereditary tumors provide evolving multicellular organisms with extra cell masses for the expression of evolutionarily novel genes and gene combinations, and thereby participate in the origin of new cell types, tissues and organs; (3) populations of tumor-bearing organisms serve as transitional forms in progressive evolution; (4) tumors may be considered as search engines for new gene combinations and morphological novelties in the space of biological possibilities.

Several non-trivial predictions of the *carcino-evo-devo* theory have been confirmed in the author's laboratory.<sup>22</sup> Furthermore, several non-trivial explanations of the *carcino-evo-devo* theory have been discussed, and the relationship between the *carcino-evo-devo* theory and other biological theories was examined.<sup>9</sup> The demonstration that human orthologs of fish *TSEEN* genes acquire progressive functions not encountered in fish was the direct confirmation of the *carcino-evo-devo* theory.<sup>25</sup>

The *carcino-evo-devo* theory has its own structure, which consists of several chapters that describe the different aspects of the theory. One of the chapters is devoted to *carcino-evo-devo* diagrams.

### Carcino-evo-devo diagrams

The *carcino-evo-devo* diagram (Fig. 3a) and its unfolded form (Fig. 3b) were suggested in a study<sup>11</sup> to describe the major postulates of the *carcino-evo-devo* theory. The *carcino-evo-devo* diagram (Fig. 3a) originated independently of the central dogma diagram (Fig. 1),

and describes the different processes. However, the diagrams had striking similarities. The shape of both diagrams was triangular, and both diagrams contained important prohibitions: the arrows, which were omitted from the diagrams (refer to the discussion on prohibitions below). The unfolded *carcino-evo-devo* diagram (Fig. 3b) suggested an idea to unfold the central dogma diagram (Fig. 1) into Figure 4. Other *carcino-evo-devo* diagrams that described evolutionarily novel tumor-like organs have been published.<sup>26</sup>

### Formula for the evolution of gene expression, the origin of new cell types, evo-devo, and complexity growth in progressive evolution

Formula in Figure 7 describes the evolution of gene expression, the origin of new cell types, *evo-devo*, and multilevel complexity growth in the course of progressive biological evolution. This consisted of four diagrams (Figs. 2, 3b, 4, and 6b), which were connected using the subset symbol  $\subset$  oriented in accordance with the diagram structure. Figure 4 describes the evolution of gene expression as related to genome evolution. Figure 6b describes the origin of new differentiated cell types in progressive evolution. Figure 3b describes the evolution of development (*evo-devo*) with hereditary tumor-bearing organisms as transitory forms (*carcino-evo-devo*). Figure 2 connects Figures 3b, 4 and 6b, and describes the relative independence and reciprocity of the processes of progressive evolutionary development at different levels of structural organization. To our knowledge, formula in Figure 7 is the first to present the compact description of the progressive evolution of living organisms at three structural levels of organization.

The initial regular diagrams (Figs. 1, 3a, and 6a) are the triangular diagrams. Figure 2 may be drawn as a triangle. All diagrams discussed in the present study contained biologically important prohibitions. The similar features of the diagrams that Figure 7 comprises may open new insights into the fundamental properties of life.

Figure 7 describes the four steps of progressive evolution, which may be extended to include the subsequent steps of complexity growth. The last arrow (Devo 5  $\rightarrow$ ) in the figure points to the future steps in evolution of development. Formula in Figure 7 presents the considerable concentration of knowledge. It describes the basic statements of the *carcino-evo-devo* theory, and the biologically significant coincidences of relatively independent events at various levels (proteome 2, differentiated 2, and Devo 2; proteome 3, differentiated 3, and Devo 3; proteome 4, differentiated 4, and Devo 4) frozen in progressive evolution (Evo1, Evo 2, Evo 3, and Evo 4) by functional feedbacks and natural selection. The frozen coincidences constitute the appearance of new progressive forms in evolution, with the expression of evolutionarily novel genes and gene combinations, new differentiated cell types, and morphological innovations. The populations of tumor-bearing organisms (Carcino 1, Carcino 2, Carcino 3, and Carcino 4) served as the intermediate transitory forms.

### The relationship with Darwinism and other evolutionary theories

The hopeful coincidences of the steps in diagrams of (Fig. 7) were subjected to natural selection. That is, the *carcino-evo-devo* theory and Figure 7 do not contradict Darwinism. The *carcino-evo-devo* theory is complementary to Darwinism, because it introduces the mechanism of complexity growth in progressive evolution, which is tumor neofunctionalization.<sup>9</sup>

Furthermore, the coincidences of the steps in diagrams of (Fig. 7) may correspond to the “punctuated equilibrium” model of evolution.<sup>27</sup>

The relationship between the *carcino-evo-devo* theory and other evolutionary theories was discussed in more detail in another article.<sup>9</sup>

### Biocomputational processes, compatibility search, and biologically significant multilevel coincidences

The author considered biological computation and compatibility search in the possibility space as mechanisms of complexity growth during progressive evolution.<sup>28</sup> The search for biologically significant multilevel coincidences may be realized through biological computational processes, with tumors as the search engines.<sup>28</sup> The starting point of formula in Figure 7 was genome 1, which was in correspondence with the central role of DNA computation in the space of unrealized biological possibilities.<sup>28</sup>

### The complexity growth in progressive evolution is a multilevel process

The consecutive steps of evolution (proteome 1, proteome 2, proteome 3, and proteome 4; differentiated 1, differentiated 2, differentiated 3, and differentiated 4; Devo 1, Devo 2, Devo 3, Devo 4, and Devo 5) describe the increase in complexity at each level. That is, the complexity growth in the progressive evolution of organisms is a multilevel process. Some of the quantitative characteristics of this process is known: 411 differentiated cell types have been described in humans,<sup>29</sup> and 479 morphological characters have been scored as an index of vertebrate morphological complexity.<sup>30</sup> These quantitative characteristics may be used in computer science to model the processes of macroevolution and its interaction with the evolution of gene expression.

### Prohibitions

As Karl Popper stated,<sup>31</sup> “Every good theory is a prohibition: it forbids certain things to happen. The more a theory forbids, the better it is.” Information flow from proteins is forbidden in diagrams in Figures 1 and 4. In Figure 2, MML cannot directly interact with MCL, but only through CL. In Figure 3, normal development (Devo) cannot directly participate in progressive evolution (Evo), but only with the aid of an intermediary transitory form (Carcino). In Figure 6, terminally differentiated cells cannot replicate, de-differentiate, or generate tumors.

Figure 7 contains more prohibitions. This shows the prohibition of the simultaneous expression of all genes that constitute the genome, especially the simultaneous expression of incompatible genes that originate during genome evolution, in a differentiated cell type. Thus, Figure 7 describes the rise and evolution of differential gene expression, which is connected with genome evolution, the origin of new differentiated cell types, and the evolution of development. The prohibitions described in Figure 7 may be ranked among the fundamental properties of life.

### The use of diagrams in biology and in other branches of science

Diagrams are widely used in biology, such as diagrams of metabolic pathways, signaling pathways, immunological pathways, and other biological pathways and networks. In these diagrams, arrows

are usually used to connect different molecular structures. The central dogma describes the information flow between different types of macromolecules within the same level of organization.

The arrows in Figures 2 and 3 connect entities that belong to different levels of organization, or describe the qualitatively different biological processes. In the earlier work conducted by the author, arrows were used to describe the interactions between different levels of organization,<sup>10</sup> the evolution of development,<sup>11</sup> and the origin of evolutionarily novel tumor-like organs.<sup>26</sup> These descriptions differed in kind from previous diagrammatic biological descriptions.

The unfolding of the central dogma diagram, which has been unaltered for more than 50 years, on the evolutionary scale, described the evolution of the genome, transcriptome and proteome (Fig. 4), demonstrating the viability of the present diagrammatic approach.

Arrows (“morphisms”) were used in the mathematical category theory, allowing these to connect “objects” of different natures. The category theory is presently being used to build new theoretical physics. This also seeks examples from other sciences that may be formalized within the framework of the category theory. The author considers that the present study may suggest that the category theory participates in building new chapters of theoretical biology.

The diagrams and formulas obtained in the present study may be used in computer science to model the processes of macroevolution.

### Diagrams and empirical data

The diagram of the central dogma of molecular biology published in 1970<sup>2</sup> was based on the previous analysis of a vast amount of empirical data on protein synthesis published by F. Crick in 1958.<sup>1</sup> The same is true for the *carcino-evo-devo* diagrams. First introduced in 2019,<sup>11</sup> the *carcino-evo-devo* diagrams were based on the theory that summarized a vast amount of empirical evidence obtained from several fields of biology. More than one thousand references have been cited in the monograph published by the author in 2014.<sup>13</sup> Cellular diagrams (Fig. 6) in the present study were based on the empirical data described in the references.<sup>11–17</sup> Thus, the diagrams in the present study represent the generalization of empirical data, and reflect the actual biological processes. The category theory uses a similar approach.

The formula in Figure 7 possesses predictive power. This predicts that at least 411 transcriptomes are necessary to determine the development of existing mammalian differentiated cell types. This emphasizes the importance of studying the newly discovered class of *TSEEN* genes in different groups of organisms, in order to understand the evolution of gene expression, including studies of epigenetic modifications. The non-trivial predictions of the *carcino-evo-devo* theory and its experimental confirmation were discussed in more detail in a reference.<sup>22</sup>

Although the central dogma mentioned nothing on the machinery of information transfers and control mechanisms, it stimulated studies of corresponding molecular mechanisms. Similarly, the *carcino-evo-devo* theory and its diagrams would stimulate studies on the molecular, cellular and multicellular mechanisms of the evolution of gene expression and complexity growth.

### Where does the information unrelated to gene sequence come from?

The author intends to respond to an important question: where

does the information that is unrelated to that in gene sequence come from?<sup>8</sup> The answer to this query would be higher levels of structural organization, as specified by Figure 7. The second significant issue that was brought up in a reference<sup>32</sup> concerned the upward vs. downward transmission of information between genotypes and phenotypes. Based on the findings of the present study, the correct response would be, as follows: in Figure 7, there is reciprocity between the processes of evolutionary development at different levels of structural organization, that is, the information is being transferred in both directions.

In conclusion, the diagram of the central dogma unfolding upwards along the evolutionary scale describes the evolution of gene expression (Fig. 4). The formula that describes the evolution of gene expression, the origin of new cell types, *evo-devo*, and the multilevel complexity growth in the course of progressive biological evolution was obtained (Fig. 7). This formula explains several fundamental questions about progressive evolution. Similar features of the diagrams in Figure 7 may open new insights into the fundamental characteristics of life. The coincidences of events at different levels described in Figure 7 may correspond to the punctuated equilibrium model of progressive evolution.

The diagrams and formulas obtained in the present study may stimulate new experimental studies of gene expression, particularly the newly discovered class of *TSEEN* genes. These diagrams and formulas may be formalized within the framework of the category theory, and used in computer science to model the interrelationship of gene expression with macroevolutionary processes.

The theory of *carcino-evo-devo* and the formula of complexity growth considers evolutionary, individual and neoplastic development at three levels of structural organization within one theoretical framework. This is the reason why the *carcino-evo-devo* theory has the potential to become a unifying biological theory.

## Acknowledgments

None.

## Funding

This study was funded by the Ministry of Science and Higher Education of the Russian Federation, under the Strategic Academic Leadership Program “Priority 2030” (Agreement 075-15-2023-380 dated 20 February 2023).

## Conflict of interest

The author declares no conflict of interest.

## References

- [1] Crick FH. On protein synthesis. *Symp Soc Exp Biol* 1958;12:138–63. PMID:13580867.
- [2] Crick F. Central dogma of molecular biology. *Nature* 1970;227(5258):561–563. doi:10.1038/227561a0, PMID:4913914.
- [3] Davidson EH (ed). *The Regulatory Genome*. Amsterdam: Elsevier Academic Press; 2006.
- [4] Mitsis T, Efthimiadou A, Bacopoulos F, Vlachakis D, Chrousos GP, Eliopoulos E. Transcription factors and evolution: An integral part of gene expression. *World Academy of Sciences Journal* 2020;2:3–8. doi:10.3892/wasj.2020.32.
- [5] Bylino OV, Ibragimov AN, Shidlovskii YV. Evolution of Regulated Transcription. *Cells* 2020;9(7):1675. doi:10.3390/cells9071675, PMID:32664620.
- [6] Mistelli T. The self-organizing genome: Principles of genome organization and function. *Cell* 2020;183:28–45. doi:10.1016/j.cell.2020.09.014, PMID:32976797.
- [7] Sood V, Misteli T. The stochastic nature of genome organization and function. *Curr Opin Genet Dev* 2022;72:45–52. doi:10.1016/j.gde.2021.10.004, PMID:34808408.
- [8] Baverstock K. Crick's sequence hypothesis - a review. *Commun Integr Biol* 2019;12(1):59–64. doi:10.1080/19420889.2019.1594501, PMID:31143364.
- [9] Kozlov AP. *Carcino-Evo-Devo, A Theory of the Evolutionary Role of Hereditary Tumors*. *Int J Mol Sci* 2023;24(10):8611. doi:10.3390/ijms24108611, PMID:37239953.
- [10] Kozlov AP. Evolution of living organisms as a multilevel process. *J Theor Biol* 1979;81(1):1–17. doi:10.1016/0022-5193(79)90076-6, PMID:529814.
- [11] Kozlov AP. The Role of Heritable Tumors in Evolution of Development: a New Theory of *Carcino-evo-devo*. *Acta Naturae* 2019;11(4):65–72. doi:10.32607/20758251-2019-11-4-65-72, PMID:31993236.
- [12] Young MD, Mitchell TJ, Vieira Braga FA, Tran MGB, Stewart BJ, Ferdinand JR, et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* 2018;361(6402):594–599. doi:10.1126/science.aat1699, PMID:30093597.
- [13] Kozlov AP (ed). *Evolution by tumor neofunctionalization. The role of tumors in the origin of new cell types, tissues and organs*. Amsterdam: Elsevier/Academic Press; 2014.
- [14] Cieřlar-Pobuda A, Knořlach V, Ringh MV, Stark J, Likus W, Siemianowicz K, et al. Transdifferentiation and reprogramming: Overview of the processes, their similarities and differences. *Biochim Biophys Acta Mol Cell Res* 2017;1864(7):1359–1369. doi:10.1016/j.bbamcr.2017.04.017, PMID:28460880.
- [15] Tanabe K, Ang CE, Chanda S, Olmos VH, Haag D, Levinson DF, et al. Transdifferentiation of human adult peripheral blood T cells into neurons. *Proc Natl Acad Sci USA* 2018;115(25):6470–6475. doi:10.1073/pnas.1720273115, PMID:29866841.
- [16] Reid A, Tursun B. Transdifferentiation: do transition states lie on the path of development? *Curr Opin Syst Biol* 2018;11:18–23. doi:10.1016/j.coisb.2018.07.004, PMID:30386832.
- [17] Trosco JE. On potential origin and characteristics of cancer stem cells. *Carcinogenesis* 2021;42(7):905–912. doi:10.1093/carcin/bgab042, PMID:34014276.
- [18] Kozlov AP (ed). *Regulatory mechanisms as an expression and the result of evolution of competitive relations between the genes. Salinity Adaptations of Aquatic Organisms*. Leningrad, USSR: Nauka; 1976:237–345.
- [19] Ohno S (ed). *Evolution by Gene Duplication*. New York: Springer-Verlag; 1970.
- [20] King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. *Science* 1975;188(4184):107–116. doi:10.1126/science.1090005, PMID:1090005.
- [21] Kozlov AP. Expression of evolutionarily novel genes in tumors. *Infect Agent Cancer* 2016;11:34. doi:10.1186/s13027-016-0077-6, PMID:27437030.
- [22] Kozlov AP. The Theory of *Carcino-Evo-Devo* and Its Non-Trivial Predictions. *Genes (Basel)* 2022;13(12):2347. doi:10.3390/genes13122347, PMID:36553613.
- [23] Kozlov AP (ed). *Evolution by Tumor Neofunctionalization: Neoplasms as the Factor of Progressive Evolution*. St. Petersburg: Publishing House of Peter the Great Polytechnic University; 2016.
- [24] Kozlov AP (ed). *Evolution by Tumor Neofunctionalization: The Role of Tumors in the Origin of New Cell Types, Tissues and Organs*. Beijing: China Science Publishing & Media Ltd (Science Press); 2019.
- [25] Matyunina EA, Emelyanov AV, Kurbatova TV, Makashov AA, Mizgirev IV, Kozlov AP. Evolutionary novel genes are expressed in transgenic fish tumors and their orthologs are involved in development of progressive traits in humans. *Infectious Agents Cancer* 2019;14:46. doi:10.1186/s13027-019-0262-5, PMID:31827597.
- [26] Kozlov AP. Mammalian tumor-like organs.1. The role of tumor-like normal organs and atypical tumor organs in the evolution of development (*carcino-evo-devo*). *Infectious Agents Cancer* 2022;17(1):2. doi:10.1186/s13027-021-00412-0, PMID:35012580.

- [27] Eldredge N, Gould SJ. Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf TJM (ed). *Models in Paleobiology*. San Francisco: Freeman Cooper & Co; 1972:82–115.
- [28] Kozlov AP. Biological computation and compatibility search in the possibility space as the mechanism of complexity increase during progressive evolution. *Evolutionary Bioinformatics* 2022;18:1–5. doi: 10.1177/11769343221110654.
- [29] Vickaryous MK, Hall BK. Human cell type diversity, evolution, development, and classification with special reference to cells derived from neural crest. *Biol Rev Camb Philo Soc* 2006;81(3):425–455. doi: 10.1017/S1464793106007068, PMID:16790079.
- [30] Aburomia R, Khaner O, Sidow A. Functional evolution in the ancestral lineage of vertebrates or when genomic complexity was wagging its morphological tail. *J Struct Func Genomics* 2003;3(1–4):45–52. PMID:12836684.
- [31] Popper K. *Conjectures and Reflection*. London: Routledge and Kegan Paul; 1963.
- [32] Annala A, Baverstock K. Genes without prominence: a reappraisal of the foundations of biology. *J R Soc Interface* 2014;11(94):20131017. doi:10.1098/rsif.2013.1017, PMID:24554573.