

Филостратиграфический анализ генной сети клеточного цикла у животных

Турнаев И.^{1, 2*}, Иванов Р.А.¹, Лашин С.А.^{1, 2, 3}, Афонников Д.А.^{1, 2, 3}

¹ Институт цитологии и генетики СО РАН, Новосибирск, Россия

² Курчатowski геномный центр ИЦиГ СО РАН, Новосибирск, Россия

³ Новосибирский государственный университет, Новосибирск, Россия

* turn@bionet.nsc.ru

Ключевые слова: генная сеть; клеточный цикл; филостратиграфический индекс; индекс возрастов; индекс дивергенции

Мотивация и цель: Клеточный цикл (КЦ) – это период существования клетки с момента ее образования путем деления материнской клетки до собственного деления или гибели. Митотический (клеточный) цикл включает четыре последовательно проходимые клеткой фазы: G1 (рост клетки, подготовка к S фазе), S (синтез ДНК), G2 (подготовка к M фазе), M (митоз – разделение материала и структур клетки на две новые клетки и их расхождение). Прохождение клеткой митотического цикла контролируется генной сетью клеточного цикла, включающей центральный транскрипционный фактор, активирующий транскрипцию значительной группы генов КЦ – E2F (E2F1-5); репрессор транскрипции RB (RB1, RBL1, RBL2); комплексы циклин/CDK, управляющие комплексом E2F/RB и через него экспрессией генов КЦ, а также контролирующие другие процессы КЦ; ингибиторы CDK (p21-CDKN1A, p27-CDKN1B и др.); белки, передающие сигналы от ростовых факторов к белкам КЦ (JUN, FOS, MAP-киназы) [1]. Чтобы лучше понять, как изменялась сеть КЦ и функции ее генов в процессе эволюции животных, мы провели анализ возрастов и изменчивости генов с учетом сетевых связей между ними.

Методы и алгоритмы: Данные о генах КЦ и их взаимодействиях были собраны из карточек генов в NCBI gene и информации в статьях PubMed. На основе собранных данных была реконструирована генная сеть клеточного цикла с помощью компьютерной системы Cytoscape [2]. Эволюционные характеристики генов (индекс филостратиграфического возраста (PAI) и индекс дивергенции (DI)) рассчитывали, как описано в [3], с учетом последовательностей ортологичных генов, идентичных рассматриваемому на 50 % и более.

Результаты: Были собраны данные о 131 гене клеточного цикла. Более половины из них (72 гена) вовлечены больше всего в клеточный цикл, 20 % (27 генов) – в транскрипцию РНК, 13 % – в репликацию ДНК, 10 % – в апоптоз, 5 % – в репарацию ДНК, 3 % – в дифференцировку клеток. На основании этой информации с помощью компьютерной системы Cytoscape была реконструирована генная сеть КЦ. Сеть включает 131 ген и 1034 взаимодействия между этими генами: 664 активирующих и 370 подавляющих. В полученной сети каждый ген в среднем взаимодействует с 13,847 генами. Наибольшее число связей с другими генами с сети КЦ имеют гены E2F1 (83 связи), P53 (72), E2F3 (70), E2F2,4,5,6 (по 68 связей), CDK2 (49), c-MYC (48).

Мы провели анализ PAI и DI у генов КЦ. Наиболее древними оказались 12 генов (9 % генов КЦ), произошедших у клеточных организмов. Происхождение 59 % генов сети КЦ соответствует появлению одноклеточных эукариот. Еще 27 % генов появилось с многоклеточными эукариотами, 1 ген – у хордовых (SMAD2), 5 генов (4 % генов КЦ) появились в эволюции у позвоночных, и появление одного гена (NPAT) соответствует появлению костных позвоночных (Euteleostomi).

Наиболее древними генами, появившимися у клеточных организмов, оказались гены, вовлеченные в ДНК репликацию, ДНК репарацию, транскрипцию, синтез фолатов, синтез нуклеотидов. Из 12 генов, появившихся у клеточных организмов, только три вовлечены в первую очередь в клеточный цикл: c-MYC, фосфатаза PP2A и SKP1. Представители семейства транскрипционных факторов E2F E2F2-6 появились в эволюции у одноклеточных эукариот, и только E2F1 появился у многоклеточных организмов. Гены семейства RB явились в эволюции с появлением одноклеточных эукариот – RB1 и RBL1, ген RBL2 – при появлении многоклеточных эукариот. Представители семейства циклинов в основном появились у одноклеточных эукариот и только три циклина (cyclins D2, E1 и E2) – у многоклеточных эукариот, а также самый молодой cyclin I появился в эволюции после появления позвоночных. Почти все гены семейства CDK (циклин-зависимых киназ) произошли у одноклеточных эукариот, за исключением CDK6 и CDK16, появившихся немного позже у многоклеточных эукариот. Интересно, что представители CDI (ингибиторы циклин-зависимых киназ) появляются в эволюции несколько позже. Только два гена CDKN2D (p19-INK4D) и CDKN3 (KAP) произошли у одноклеточных эукариот, гены CDKN1A, CDKN1B, CDKN1C, CDKN2C произошли у многоклеточных эукариот, и два гена CDKN2A (ARF), CDKN2B (p15INK4b) произошли совместно с формированием в эволюции позвоночных.

Все молодые гены КЦ, произошедшие начиная с хордовых (хордовые, позвоночные и косные позвоночные), отличаются низким уровнем дивергенции генов DI (<0.3). Это указывает на то, что данные гены находятся под сильным селективным отбором.

Выводы: Проанализировав публикации PubMed, мы собрали данные по 131 гену клеточного цикла. Анализ возрастов (индекс PAI) генов сети КЦ показал, что представители основных семейств генов КЦ E2F, RB, CDK и циклинов появились в эволюции совместно с образованием одноклеточных или многоклеточных эукариот. Тогда как гены семейства CDI появились несколько позже в основном у многоклеточных эукариот и у позвоночных. Это может быть связано с необходимостью более жесткого контроля КЦ у многоклеточных эукариот и позвоночных.

Финансирование: Исследование поддержано бюджетным проектом № FWNR-2022-0020.

Phylostratigraphic analysis of gene networks of cell cycle in animals

Turnaev I.I.^{1, 2*}, Ivanov R.A.¹, Lashin S.A.^{1, 2, 3}, Afonnikov D.A.^{1, 2, 3}

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

² Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

³ Novosibirsk State University, Novosibirsk, Russia

* turn@bionet.nsc.ru

Key words: gene network; cell cycle; phylostratigraphic age index; divergence index

Motivation and Aim: The cell cycle (CC) is the period of a cell's existence from the moment of its formation by division of the mother cell to its own division or death. The mitotic (cellular) cycle includes four phases that the cell passes through sequentially: G1 (cell growth, preparation for the S phase), S (DNA synthesis), G2 (preparation for the M phase), M (mitosis – separation of cell material and structures into two new cells and their divergence). The mitotic cycle of the cell is controlled by the cell cycle gene network, including the central transcription factor activating the transcription of a significant group of CC genes – E2F (E2F1-5); Rb family (RB1, RBL1, RBL2) are best known as a repressors of the E2F/DP family of transcription factors; cyclin/CDK complexes that control the E2F/RB complex and through it the expression of CC genes, as well as control other CC processes; CDK inhibitors (p21-CDKN1A, p27-CDKN1B and etc.); proteins (JUN, FOS, MAP-kinases) that transmit signals from growth factors to CC proteins [1]. To better understand how the CC network and the function of its genes have changed in evolution, we analyzed the ages and variability of genes with respect to the network connections between them.

Methods and Algorithms: Data on CC genes and their interactions were collected from gene cards in NCBI's Gene and information in PubMed articles. Based on the collected data, the gene network of the cell cycle was reconstructed using the Cytoscape computer system [2]. The evolutionary characteristics of genes ((phylostratigraphic age index (PAI) and divergence index (DI)) were calculated as described in [3], taking into account the sequences of orthologous genes identical to the considered gene by 50 % or more.

Results: Data on 131 cell cycle genes were collected. More than half of them (72 genes) are most involved in the cell cycle, 20 % (27 genes) in RNA transcription, 13 % in DNA replication, 10 % in apoptosis, 5 % in DNA repair and 3 % in cell differentiation. Based on this information, the CC gene network was reconstructed using the Cytoscape computer system. The network includes 131 genes and 1034 interactions between these genes: 664 activating and 370 repressing. In the resulting network, each gene interacts with 13,847 genes on average. E2F1 (83 interactions), P53 (72), E2F3 (70), E2F2,4,5,6 (68 interactions each), CDK2 (49), c-MYC (48) genes have the highest number of interactions with other genes from the CC network.

We analyzed PAI and DI in CC genes. The most ancient were 12 genes (9 % of CC genes) originated in cellular organisms. The origin of 59 % of the genes of the CC network corresponds to the appearance of unicellular eukaryotes. Another 27 % of genes appeared with multicellular eukaryotes, one gene appeared in chordates (SMAD2), 5 genes (4 % of CC genes) appeared in evolution in vertebrates and the appearance of 1 gene (NPAT) corresponds to the appearance of bony vertebrates (Euteleostomi).

The most ancient genes that appeared in cellular organisms were genes involved in DNA replication, DNA repair, transcription, folate synthesis, and nucleotide synthesis. Of the 12 genes that have appeared in cellular organisms, only three are primarily involved in the cell cycle: c-MYC, phosphatase PP2A, and SKP1. Genes of the E2F family of transcription factors E2F2-6 appeared in evolution in unicellular eukaryotes and only E2F1 appeared in multicellular eukaryotes. Genes of the RB family appeared in evolution with the appearance of unicellular eukaryotes – RB1, RBL1, but the RBL2 gene – at the appearance of multicellular eukaryotes. Genes of the cyclin family mainly appeared in unicellular eukaryotes and only 3 cyclins (cyclins D2, E1 and E2) appeared in multicellular eukaryotes, as well as the youngest cyclin I appeared in evolution after

the appearance of vertebrates. Almost all genes of the CDK (cyclin-dependent kinases) family originated in unicellular eukaryotes, with the exception of CDK6 and CDK16, which appeared a little later, in multicellular eukaryotes. Interestingly, genes of CDIs (inhibitors of cyclin-dependent kinases) appear somewhat later in evolution. Only two genes CDKN2D (p19-INK4D) and CDKN3 (KAP) occurred in unicellular eukaryotes, genes CDKN1A, CDKN1B, CDKN1C, CDKN2C occurred in multicellular eukaryotes and two genes CDKN2A (ARF), CDKN2B (p15-INK4b) occurred together with formation in the evolution of vertebrates.

All young CC genes from chordates (chordates, vertebrates, and bony vertebrates) are characterized by a low level of DI gene divergence (< 0.3). This indicates that these genes are under high selective pressure.

Conclusion: Having analyzed Pubmed publications, we collected data on 131 cell cycle genes. Analysis of the ages (PAI index) of the genes of the CC network showed that representatives of the major CC gene families E2F, RB, CDK, and cyclins appeared in evolution together with the formation of unicellular or multicellular eukaryotes. Whereas CDI family genes appeared somewhat later mainly in multicellular eukaryotes and in vertebrates. This may be due to the need for tighter control of CC in multicellular eukaryotes and vertebrates.

Funding: The study is supported by the budget project No. FWNR-2022-0020.

Список литературы/References

1. Johnson D., Walker C. Cyclins and cell cycle checkpoints. *Annu Rev Pharmacol Toxicol.* 1999;39:295-312. doi 10.1146/annurev.pharmtox.39.1.295
2. Shannon P., Markiel A., Ozier O. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-2504
3. Mustafin Z.S. et al. Phylostratigraphic analysis of gene networks of human diseases. *Vavilov J Genet Breed.* 2021;25(1):46-56