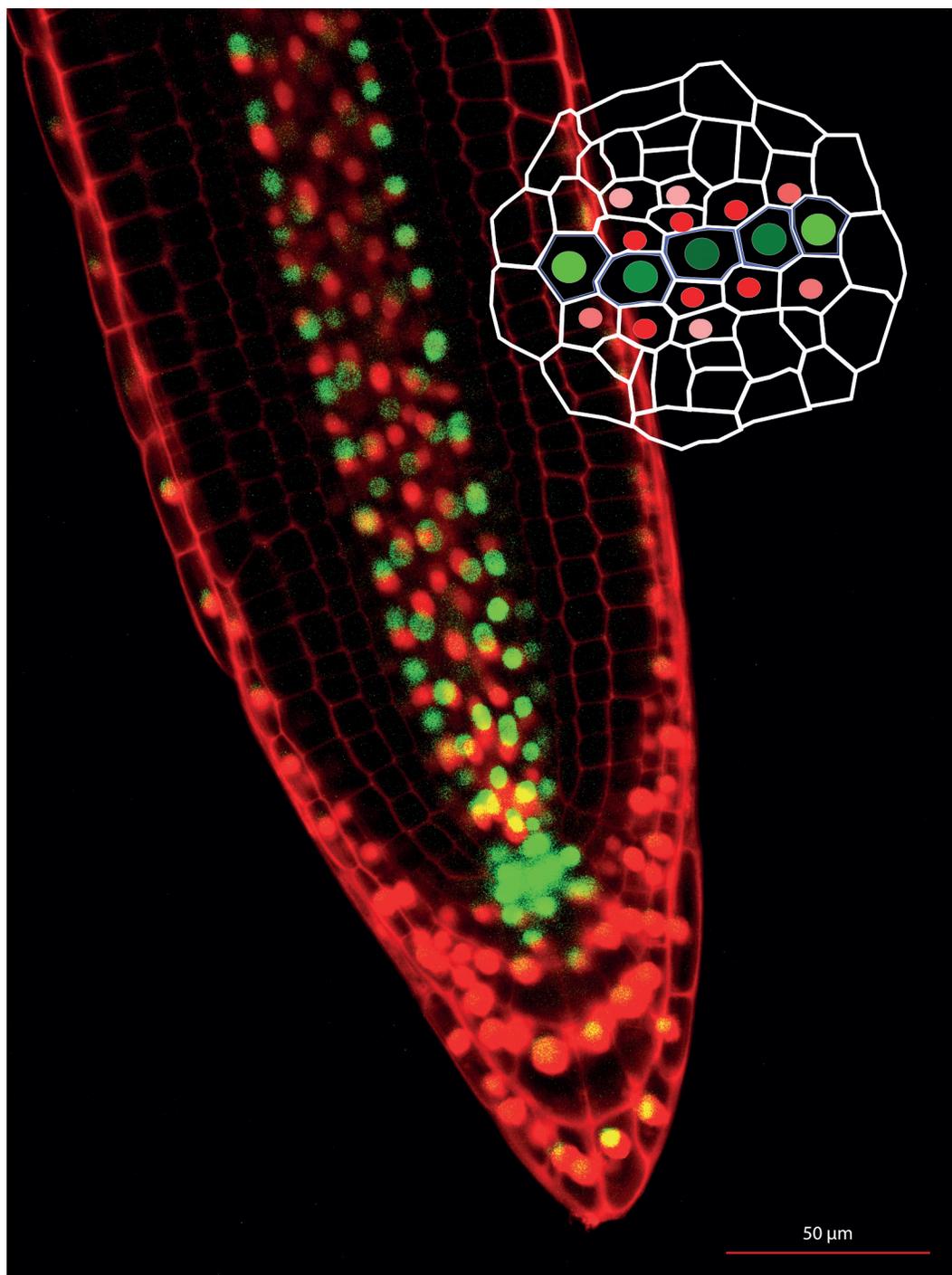


# ВАВИЛОВСКИЙ

## ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

VAVILOV JOURNAL OF GENETICS AND BREEDING

2022 • 26 • 8



Системная компьютерная биология / Эволюционная компьютерная биология / Компьютерная геномика

vavilov.ejpub.ru/  
bioenet.nsc.ru/vogis  
vavilov\_journal@bioenet.nsc.ru  
Индекс издания 42153





Научный рецензируемый журнал

# ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

Основан в 1997 г.

Периодичность 8 выпусков в год

DOI 10.18699/VJGB-22-86

## Учредители

Сибирское отделение Российской академии наук

Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук»

Межрегиональная общественная организация Вавиловское общество генетиков и селекционеров

## Главный редактор

*А.В. Кочетов* – академик РАН, д-р биол. наук (Россия)

## Заместители главного редактора

*Н.А. Колчанов* – академик РАН, д-р биол. наук, профессор (Россия)

*И.Н. Леонова* – д-р биол. наук (Россия)

*Н.Б. Рубцов* – д-р биол. наук, профессор (Россия)

*В.К. Шумный* – академик РАН, д-р биол. наук, профессор (Россия)

## Ответственный секретарь

*Г.В. Орлова* – канд. биол. наук (Россия)

## Редакционная коллегия

*Е.Е. Андронов* – канд. биол. наук (Россия)

*Ю.С. Аульченко* – д-р биол. наук (Россия)

*О.С. Афанасенко* – академик РАН, д-р биол. наук (Россия)

*Д.А. Афонников* – канд. биол. наук, доцент (Россия)

*Л.И. Афтанас* – академик РАН, д-р мед. наук (Россия)

*Л.А. Беспалова* – академик РАН, д-р с.-х. наук (Россия)

*А. Бёрнер* – д-р наук (Германия)

*Н.П. Бондарь* – канд. биол. наук (Россия)

*С.А. Боринская* – д-р биол. наук (Россия)

*П.М. Бородин* – д-р биол. наук, проф. (Россия)

*А.В. Васильев* – чл.-кор. РАН, д-р биол. наук (Россия)

*М.И. Воевода* – академик РАН, д-р мед. наук (Россия)

*Т.А. Гавриленко* – д-р биол. наук (Россия)

*И. Гроссе* – д-р наук, проф. (Германия)

*Н.Е. Грунтенко* – д-р биол. наук (Россия)

*С.А. Демаков* – д-р биол. наук (Россия)

*И.К. Захаров* – д-р биол. наук, проф. (Россия)

*И.А. Захаров-Гезехус* – чл.-кор. РАН, д-р биол. наук (Россия)

*С.Г. Инге-Вечтомов* – академик РАН, д-р биол. наук (Россия)

*А.В. Кильчевский* – чл.-кор. НАНБ, д-р биол. наук (Беларусь)

*С.В. Костров* – чл.-кор. РАН, д-р хим. наук (Россия)

*А.М. Кудрявцев* – чл.-кор. РАН, д-р биол. наук (Россия)

*И.Н. Лаврик* – д-р биол. наук (Германия)

*Д.М. Ларкин* – канд. биол. наук (Великобритания)

*Ж. Ле Гуи* – д-р наук (Франция)

*И.Н. Лебедев* – д-р биол. наук, проф. (Россия)

*Л.А. Лутова* – д-р биол. наук, проф. (Россия)

*Б. Люгтенберг* – д-р наук, проф. (Нидерланды)

*В.Ю. Макеев* – чл.-кор. РАН, д-р физ.-мат. наук (Россия)

*В.И. Молодин* – академик РАН, д-р ист. наук (Россия)

*М.П. Мошкин* – д-р биол. наук, проф. (Россия)

*С.Р. Мурсалимов* – канд. биол. наук (Россия)

*Л.Ю. Новикова* – д-р с.-х. наук (Россия)

*Е.К. Потокина* – д-р биол. наук (Россия)

*В.П. Пузырев* – академик РАН, д-р мед. наук (Россия)

*Д.В. Пышный* – чл.-кор. РАН, д-р хим. наук (Россия)

*И.Б. Розозин* – канд. биол. наук (США)

*А.О. Рувинский* – д-р биол. наук, проф. (Австралия)

*Е.Ю. Рыкова* – д-р биол. наук (Россия)

*Е.А. Салина* – д-р биол. наук, проф. (Россия)

*В.А. Степанов* – академик РАН, д-р биол. наук (Россия)

*И.А. Тихонович* – академик РАН, д-р биол. наук (Россия)

*Е.К. Хлесткина* – д-р биол. наук, проф. РАН (Россия)

*Э.К. Хуснутдинова* – д-р биол. наук, проф. (Россия)

*М. Чен* – д-р биол. наук (Китайская Народная Республика)

*Ю.Н. Шавруков* – д-р биол. наук (Австралия)

*Р.И. Шейко* – чл.-кор. НАНБ, д-р с.-х. наук (Беларусь)

*С.В. Шестаков* – академик РАН, д-р биол. наук (Россия)

*Н.К. Янковский* – академик РАН, д-р биол. наук (Россия)

Scientific Peer Reviewed Journal

# VAVILOV JOURNAL OF GENETICS AND BREEDING

## VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII

*Founded in 1997**Published 8 times annually*

DOI 10.18699/VJGB-22-86

**Founders**

Siberian Branch of the Russian Academy of Sciences

Federal Research Center Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences

The Vavilov Society of Geneticists and Breeders

**Editor-in-Chief**

A.V. Kochetov, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia

**Deputy Editor-in-Chief**

N.A. Kolchanov, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia

I.N. Leonova, Dr. Sci. (Biology), Russia

N.B. Rubtsov, Professor, Dr. Sci. (Biology), Russia

V.K. Shumny, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia

**Executive Secretary**

G.V. Orlova, Cand. Sci. (Biology), Russia

**Editorial board**

O.S. Afanasenko, Full Member of the RAS, Dr. Sci. (Biology), Russia

D.A. Afonnikov, Associate Professor, Cand. Sci. (Biology), Russia

L.I. Aftanas, Full Member of the RAS, Dr. Sci. (Medicine), Russia

E.E. Andronov, Cand. Sci. (Biology), Russia

Yu.S. Aulchenko, Dr. Sci. (Biology), Russia

L.A. Bespalova, Full Member of the RAS, Dr. Sci. (Agricul.), Russia

N.P. Bondar, Cand. Sci. (Biology), Russia

S.A. Borinskaya, Dr. Sci. (Biology), Russia

P.M. Borodin, Professor, Dr. Sci. (Biology), Russia

A. Börner, Dr. Sci., Germany

M. Chen, Dr. Sci. (Biology), People's Republic of China

S.A. Demakov, Dr. Sci. (Biology), Russia

T.A. Gavrilenko, Dr. Sci. (Biology), Russia

I. Grosse, Professor, Dr. Sci., Germany

N.E. Gruntenko, Dr. Sci. (Biology), Russia

S.G. Inge-Vechtomov, Full Member of the RAS, Dr. Sci. (Biology), Russia

E.K. Khlestkina, Professor of the RAS, Dr. Sci. (Biology), Russia

E.K. Khusnutdinova, Professor, Dr. Sci. (Biology), Russia

A.V. Kilchevsky, Corr. Member of the NAS of Belarus, Dr. Sci. (Biology),

Belarus

S.V. Kostrov, Corr. Member of the RAS, Dr. Sci. (Chemistry), Russia

A.M. Kudryavtsev, Corr. Member of the RAS, Dr. Sci. (Biology), Russia

D.M. Larkin, Cand. Sci. (Biology), Great Britain

I.N. Lavrik, Dr. Sci. (Biology), Germany

J. Le Gouis, Dr. Sci., France

I.N. Lebedev, Professor, Dr. Sci. (Biology), Russia

B. Lugtenberg, Professor, Dr. Sci., Netherlands

L.A. Lutova, Professor, Dr. Sci. (Biology), Russia

V.Yu. Makeev, Corr. Member of the RAS, Dr. Sci. (Physics and Mathem.),

Russia

V.I. Molodin, Full Member of the RAS, Dr. Sci. (History), Russia

M.P. Moshkin, Professor, Dr. Sci. (Biology), Russia

S.R. Mursalimov, Cand. Sci. (Biology), Russia

L.Yu. Novikova, Dr. Sci. (Agricul.), Russia

E.K. Potokina, Dr. Sci. (Biology), Russia

V.P. Puzyrev, Full Member of the RAS, Dr. Sci. (Medicine),  
RussiaD.V. Pyshnyi, Corr. Member of the RAS, Dr. Sci. (Chemistry),  
Russia

I.B. Rogozin, Cand. Sci. (Biology), United States

A.O. Ruvinsky, Professor, Dr. Sci. (Biology), Australia

E.Y. Rykova, Dr. Sci. (Biology), Russia

E.A. Salina, Professor, Dr. Sci. (Biology), Russia

Y.N. Shavrukov, Dr. Sci. (Biology), Australia

R.I. Sheiko, Corr. Member of the NAS of Belarus,  
Dr. Sci. (Agricul.), BelarusS.V. Shestakov, Full Member of the RAS, Dr. Sci. (Biology),  
RussiaV.A. Stepanov, Full Member of the RAS, Dr. Sci. (Biology),  
RussiaI.A. Tikhonovich, Full Member of the RAS, Dr. Sci. (Biology),  
Russia

A.V. Vasiliev, Corr. Member of the RAS, Dr. Sci. (Biology), Russia

M.I. Voevoda, Full Member of the RAS, Dr. Sci. (Medicine),  
RussiaN.K. Yankovsky, Full Member of the RAS, Dr. Sci. (Biology),  
Russia

I.K. Zakharov, Professor, Dr. Sci. (Biology), Russia

I.A. Zakharov-Gezekhus, Corr. Member of the RAS,  
Dr. Sci. (Biology), Russia

719

ОТ РЕДАКТОРА

Н.А. Колчанов, Ю.Г. Матушкин

## Системная компьютерная биология

721

ОБЗОР

Молекулярные механизмы детерминации клеток сосудистой системы корня *Arabidopsis thaliana* L.

А.Д. Сидоренко, Н.А. Омелянчук, Е.В. Землянская

733

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Компьютерный анализ особенностей регуляции гиперметилированных маркерных генов гепатокарциномы вирусными белками гепатита С.

Е.А. Антропова, Т.М. Хлебодарова, П.С. Деменков, А.С. Вензель, Н.В. Иванисенко, А.Д. Гавриленко, Т.В. Иванисенко, А.В. Адамовская, П.М. Ревва, И.Н. Лаврик, В.А. Иванисенко

743

ОБЗОР

Рациональная метаболическая инженерия *Corynebacterium glutamicum* для продукции L-валина. М.Е. Шереметьева, К.Э. Ануфриев, Т.М. Хлебодарова, Н.А. Колчанов, А.С. Яненко

758

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Стратификация и слоения в фазовых портретах моделей генных сетей.

В.П. Голубятников, А.А. Акиншин, Н.Б. Аюпова, Л.С. Минушкина

765

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Апробация технологии оценки мимики лиц для изучения динамики функциональных состояний человека в ЭЭГ-парадигме покоя. А.Н. Савостьянов, Е.Г. Вергунов, А.Е. Сапрыгин, Д.А. Лебедкин

773

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Разработка нейронной сети для диагностики риска возникновения депрессии по экспериментальным данным стоп-сигнал парадигмы. М.О. Зеленских, А.Е. Сапрыгин, С.С. Таможников, П.Д. Рудыч, Д.А. Лебедкин, А.Н. Савостьянов

## Эволюционная компьютерная биология

780

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Программная система на основе 3D симулятора для моделирования эволюции в популяции организмов, обладающих зрительной системой.

А.П. Девятериков, А.Ю. Пальянов

787

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Фосфолипазы А2 человека: функциональный и эволюционный анализ. И.И. Турнаев, М.Е. Бочарникова, Д.А. Афонников

798

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Промоторы генов, кодирующих β-амилазу, альбумин и глобулин пищевых растений в сравнении с непищевыми, характеризуются более низкой аффинностью к ТАТА-связывающему белку: *in silico* анализ. О.В. Вишневецкий, И.В. Чадаева, Е.Б. Шарыпова, Б.М. Хандаев, К.А. Золотарева, А.В. Казачек, П.М. Пономаренко, Н.Л. Подколонный, Д.А. Рассказов, А.Г. Богомолов, О.А. Подколонная, Л.К. Савинкова, Е.В. Землянская, М.П. Пономаренко

## Компьютерная геномика

806

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

FastContext: инструмент для контекстного анализа последовательностей в данных секвенирования нового поколения (NGS). Э. Весна, В.С. Фишман (на англ. языке)

810

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Сравнительный генетический анализ O-антигенов бактерий семейства Oxalobacteraceae: уникальность или тривиальность? С.Д. Афонникова, А.С. Комиссаров, П.Д. Кучур (на англ. языке)

819

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Контекстные сигналы в митохондриальных микроРНК млекопитающих. О.В. Вишневецкий, П.С. Ворожейкин, И.И. Титов

826

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Свойства малого мира научных организаций определяют динамику публикационной активности в области мРНК. А.Б. Фирсов, И.И. Титов

830

Алфавитный указатель авторов статей, опубликованных в журнале в 2022 г.

- 719 **FROM THE EDITOR**  
*N.A. Kolchanov, Yu.G. Matushkin*
- Systems computational biology**
- 721 **REVIEW**  
Molecular mechanisms of vascular tissue patterning in *Arabidopsis thaliana* L. roots.  
*A.D. Sidorenko, N.A. Omelyanchuk, E.V. Zemlyanskaya*
- 733 **ORIGINAL ARTICLE**  
Computer analysis of regulation of hepatocarcinoma marker genes hypermethylated by HCV proteins.  
*E.A. Antropova, T.M. Khlebodarova, P.S. Demenkov, A.S. Venzel, N.V. Ivanisenko, A.D. Gavrilenko, T.V. Ivanisenko, A.V. Adamovskaya, P.M. Revva, I.N. Lavrik, V.A. Ivanisenko*
- 743 **REVIEW**  
Rational metabolic engineering of *Corynebacterium glutamicum* to create a producer of L-valine. *M.E. Sheremetieva, K.E. Anufriev, T.M. Khlebodarova, N.A. Kolchanov, A.S. Yanenko*
- 758 **ORIGINAL ARTICLE**  
Stratifications and foliations in phase portraits of gene network models.  
*V.P. Golubyatnikov, A.A. Akinshin, N.B. Ayupova, L.S. Minushkina*
- 765 **ORIGINAL ARTICLE**  
Validation of a face image assessment technology to study the dynamics of human functional states in the EEG resting-state paradigm. *A.N. Savostyanov, E.G. Vergunov, A.E. Saprygin, D.A. Lebedkin*
- 773 **ORIGINAL ARTICLE**  
Development of a neural network for diagnosing the risk of depression according to the experimental data of the stop signal paradigm. *M.O. Zelenskih, A.E. Saprygin, S.S. Tamozhnikov, P.D. Rudych, D.A. Lebedkin, A.N. Savostyanov*

## Evolutionary computational biology

- 780 **ORIGINAL ARTICLE**  
A software system for modeling evolution in a population of organisms with vision, interacting with each other in 3D simulator.  
*A.P. Devyaterikov, A.Yu. Palyanov*
- 787 **ORIGINAL ARTICLE**  
Human phospholipases A2: a functional and evolutionary analysis.  
*I.I. Turnaev, M.E. Bocharnikova, D.A. Afonnikov*
- 798 **ORIGINAL ARTICLE**  
Promoters of genes encoding  $\beta$ -amylase, albumin and globulin in food plants have weaker affinity for TATA-binding protein as compared to non-food plants: *in silico* analysis. *O.V. Vishnevsky, I.V. Chadaeva, E.B. Sharypova, B.M. Khandaev, K.A. Zolotareva, A.V. Kazachek, P.M. Ponomarenko, N.L. Podkolodny, D.A. Rasskazov, A.G. Bogomolov, O.A. Podkolodnaya, L.K. Savinkova, E.V. Zemlyanskaya, M.P. Ponomarenko*

## Computational genomics

- 806 **ORIGINAL ARTICLE**  
FastContext: A tool for identification of adapters and other sequence patterns in next generation sequencing (NGS) data.  
*E. Viesná, V. Fishman*
- 810 **ORIGINAL ARTICLE**  
Unique or not unique? Comparative genetic analysis of bacterial O-antigens from the Oxalobacteraceae family. *S.D. Afonnikova, A.S. Komissarov, P.D. Kuchur*
- 819 **ORIGINAL ARTICLE**  
The context signals of mitochondrial miRNAs (mitomiRs) of mammals.  
*O.V. Vishnevsky, P.S. Vorozheykin, I.I. Titov*
- 826 **ORIGINAL ARTICLE**  
Small world of the miRNA science drives its publication dynamics.  
*A.B. Firsov, I.I. Titov*
- 830 Alphabetical author index for the list of papers published in the journal in 2022



Н.А. Колчанов



Ю.Г. Матушкин

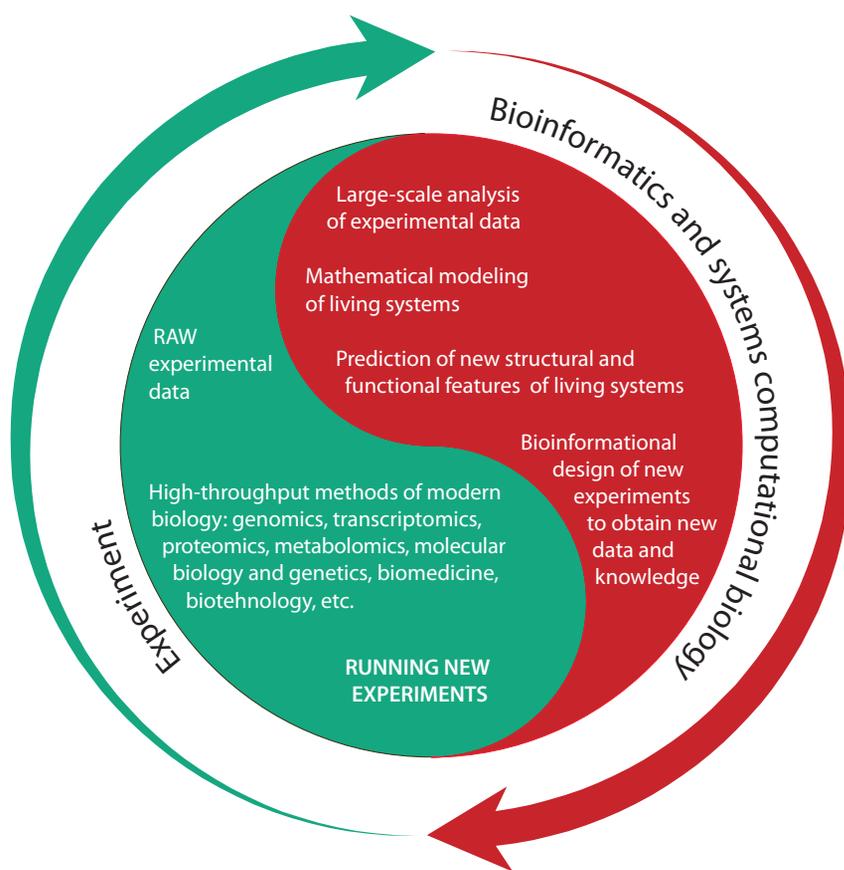
В последние 15 лет в генетике происходит информационный взрыв, обусловленный появлением эффективных методов расшифровки геномов и, как следствие, снижением более чем на четыре порядка стоимости геномного секвенирования. В результате этого науки о жизни стали главным источником больших данных, обогнав по темпам роста не только другие научные направления, но даже социальные сети. Одновременно совершенствовались экспериментальные методы транскриптомного, протеомного, метаболомного и других видов исследований. Все это создало грандиозный вызов для биоинформатики и системной компьютерной биологии, поскольку темпы генерации геномных и других типов данных намного опережают возможности их глубокого компьютерного анализа.

Объем и сложность этих данных настолько велики, что их понимание, интерпретация и, тем более, практическое применение невозможны без использования новых информационных технологий, эффективных методов анализа данных и компьютерного моделирования живых систем. На наших глазах возникает новая биология, ключевой особенностью которой является тесная интеграция экспериментальных и компьютерных подходов. Она включает:

- накопление больших объемов экспериментальных данных о структурной и функциональной организации живых систем, процессов и структур;
- крупномасштабный анализ получаемой экспериментальной информации в контексте накопленных ранее больших данных (молекулярно-биологических, генетических, биохимических, физиологических и др.);
- компьютерное моделирование изучаемых систем, процессов и структур на различных иерархических уровнях их организации;
- предсказание новых свойств и изучаемых живых систем на основе результатов анализа и моделирования;
- планирование новых экспериментов для подтверждения результатов предсказаний и прогнозов и, наконец, проведение новых экспериментов и получение новых данных и знаний.

И вот таким образом стремительно развивается новая биология, движущей силой которой является интеграция экспериментальных и компьютерных подходов. В этой интеграции важнейшую роль играют:

- а) биоинформатика, обеспечивающая хранение, обработку и анализ больших данных, получаемых с помощью методов геномики, транскриптомики, протеомики, метаболомики и других высокопроизводительных экспериментальных технологий, а также
- б) системная компьютерная биология, к числу ключевых задач которой относятся реконструкция, компьютерный анализ и моделирование генных сетей и молекулярно-генетических систем, ответственных за контроль процессов формирования молекулярно-генетических, биохимических, физиологических, структурно-морфологических, поведенческих и других характеристик человека, животных, растений и микроорганизмов на основе информации, закодированной в их геномах.



При этом биоинформатика и системная компьютерная биология имеют важнейшее значение не только для проведения фундаментальных исследований, но и для решения практических задач в интересах сельского хозяйства, биомедицины, фармакологии, биотехнологий, экологии и множества других направлений в науках о жизни и междисциплинарных исследованиях.

В очередной выпуск «Вавиловского журнала генетики и селекции» включены статьи, подготовленные по материалам ряда докладов, представленных на 13-й между-

народной мультikonференции «Биоинформатика регуляции геномов/системная биология» (4–8 июля 2022 г.) по таким направлениям, как компьютерный анализ бактериальных и митохондриальных геномов, компьютерная вирусология (взаимодействия в системе «вирус–хозяин»), рациональная инженерия метаболических путей бактерий, моделирование динамики генных сетей и процессы морфогенеза растений, эволюционная компьютерная биология, экспериментально-компьютерная психология и ряд других.

*Научные редакторы выпуска:*

*академик Н.А. Колчанов,  
научный руководитель ФИЦ ИЦиГ СО РАН*

*канд. биол. наук Ю.Г. Матушкин,  
вед. науч. сотрудник ФИЦ ИЦиГ СО РАН*

Перевод на английский язык <https://vavilov.elpub.ru/jour>

## Молекулярные механизмы детерминации клеток сосудистой системы корня *Arabidopsis thaliana* L.

А.Д. Сидоренко<sup>1,2</sup>, Н.А. Омелянчук<sup>1</sup>, Е.В. Землянская<sup>1,2</sup> 

<sup>1</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>2</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

 ezemlyanskaya@bionet.nsc.ru

**Аннотация.** Сосудистая система является результатом ароморфоза, который позволил растениям успешно освоить сушу. За счет нее осуществляется проведение воды, минеральных и органических соединений, обеспечивается эффективное сообщение между органами, а также выполняется функция механической опоры. Процесс формирования сосудистой системы – общепринятый объект фундаментальных исследований в области биологии развития растений. В частности, ранние этапы развития сосудистой системы корня модельного растения *Arabidopsis thaliana* представляют собой яркий пример самоорганизации бисимметричного (имеющего две плоскости симметрии) паттерна распределения фитогормонов, который направляет детерминацию клеток сосудистой системы. В процессе формирования сосудистой системы корня можно выделить четыре этапа: 1) детерминацию (спецификацию) клеток-предшественников проваскулярной меристемы на ранних стадиях эмбриогенеза; 2) рост и разметку проваскулярной меристемы зародыша; 3) постэмбриональное поддержание инициалей (стволовых клеток) сосудистой системы в апикальной меристеме корня; 4) конечную специализацию (дифференцировку) их дочерних клеток. Анатомические детали развития сосудистой системы *A. thaliana* давно известны и подробно описаны, однако наши знания о молекулярно-генетических механизмах этого процесса все еще ограничены. В последние годы сделано несколько важных открытий, проливающих свет на регуляцию самых ранних событий, предшествующих дифференцировке клеток сосудистой системы. В настоящем обзоре мы обобщаем данные о молекулярно-генетических механизмах, определяющих направление клеточной дифференцировки в элементы сосудистой системы корня у *A. thaliana*. Первая часть обзора посвящена описанию гистогенеза сосудистой системы корня. Далее мы реконструируем последовательность регуляторных событий, которые лежат в основе этого гистогенеза и обуславливают развитие предшественников инициалей сосудистой системы у зародыша и организацию инициалей сосудистой системы в корне проростка.

Ключевые слова: меристема; ксилема; флоэма; (про)камбий; фитогормоны; ауксин; цитокинин; *Arabidopsis thaliana*.

**Для цитирования:** Сидоренко А.Д., Омелянчук Н.А., Землянская Е.В. Молекулярные механизмы детерминации клеток сосудистой системы корня *Arabidopsis thaliana* L. *Вавиловский журнал генетики и селекции*. 2022;26(8): 721-732. DOI 10.18699/VJGB-22-88

## Molecular mechanisms of vascular tissue patterning in *Arabidopsis thaliana* L. roots

A.D. Sidorenko<sup>1,2</sup>, N.A. Omelyanchuk<sup>1</sup>, E.V. Zemlyanskaya<sup>1,2</sup> 

<sup>1</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

 ezemlyanskaya@bionet.nsc.ru

**Abstract.** A vascular system in plants is a product of aromorphosis that enabled them to colonize land because it delivers water, mineral and organic compounds to plant organs and provides effective communications between organs and mechanical support. Vascular system development is a common object of fundamental research in plant development biology. In the model plant *Arabidopsis thaliana*, early stages of vascular tissue formation in the root are a bright example of the self-organization of a bisymmetric (having two planes of symmetry) pattern of hormone distribution, which determines vascular cell fates. In the root, vascular tissue development comprises four stages: (1) specification of progenitor cells for the provascular meristem in early embryonic stages, (2) the growth and patterning of the embryo provascular meristem, (3) postembryonic maintenance of the cell identity in the vascular tissue initials within the root apical meristem, and (4) differentiation of their descendants. Although the anatomical details of *A. thaliana* root vasculature development have long been known and described in detail, our knowledge of the underlying molecular and genetic mechanisms remains limited. In recent years, several important advances have been made, shedding light on the regulation of the earliest events in provascular cells specification. In this review, we summarize the latest data on the molecular and genetic mechanisms of vascular tissue patterning in *A. thaliana*

root. The first part of the review describes the root vasculature ontogeny, and the second reconstructs the sequence of regulatory events that underlie this histogenesis and determine the development of the progenitors of the vascular initials in the embryo and organization of vascular initials in the seedling root.

Key words: meristem; xylem; phloem; (pro)cambium; plant hormones; auxin; cytokinin; *Arabidopsis thaliana*.

**For citation:** Sidorenko A.D., Omelyanchuk N.A., Zemlyanskaya E.V. Molecular mechanisms of vascular tissue patterning in *Arabidopsis thaliana* L. roots. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):721-732. DOI 10.18699/VJGB-22-88

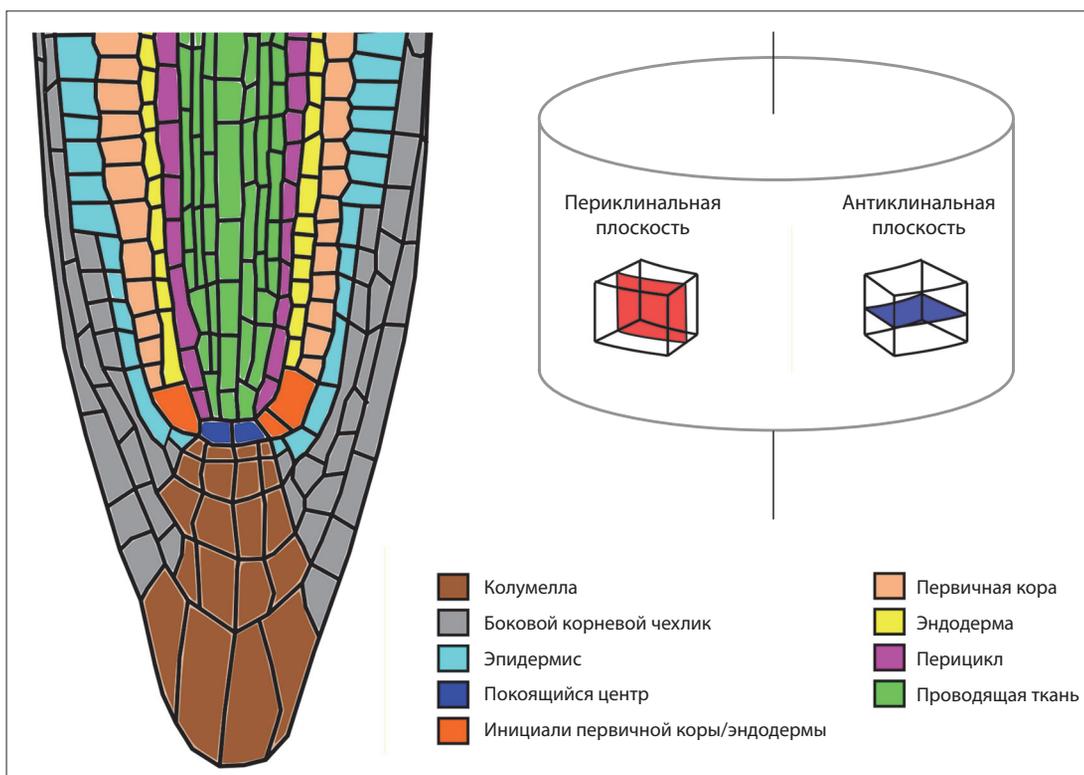
## Введение

В эволюции возникновение сосудистой системы являлось необходимой предпосылкой для выхода растений на сушу (Lucas et al., 2013). Наличие механических элементов в сосудистой системе и эффективность транспорта по ней воды, минеральных и органических веществ, а также сигнальных молекул позволили растениям достигать рекордных размеров и заселять различные территории. Сосудистая система включает в себя два различных структурно-функциональных домена: ксилему, которая осуществляет транспорт воды и минеральных веществ от корня к наземным органам, и флоэму, которая проводит органические вещества от фотосинтезирующих частей растения по направлению к корню (Evert, Eichhorn, 2006).

У покрытосеменных зрелая ксилема состоит из проводящих элементов (сосудов), которые осуществляют транспорт воды, волокон, основной функцией которых является механическая поддержка растения, и клеток паренхимы (Evert, Eichhorn, 2006). Сосуды представляют собой полые трубки, образованные соединенными в ряд клетками

с перфорациями в **антиклинальных** стенках и порами в **периклинальных** стенках (рис. 1). Сосуды и волокна формируются в результате запрограммированной гибели клеток, сформировавших лигнифицированную вторичную клеточную стенку (Courtois-Moreau et al., 2009; Smith et al., 2013; Furuta et al., 2014). Живые клетки паренхимы выполняют запасающую функцию, а также участвуют в лигнификации сосудов и регуляции скорости транспорта воды (Ménard, Pesquet, 2015; Růžicka et al., 2015).

Флоэма покрытосеменных образована ситовидными трубками, осуществляющими транспорт пластических веществ; клетками-спутницами; волокнами и склереидами, обеспечивающими механическую поддержку; а также клетками паренхимы (Sjolund, 1997; Evert, Eichhorn, 2006). В отличие от полых лигнифицированных сосудов ксилемы, ситовидные трубки представляют собой тяж живых клеток (ситовидных элементов), которые сообщаются ситовидными полями – скоплением мелких пор в антиклинальных стенках. Ситовидные элементы формируют утолщенную нелигнифицированную вторичную клеточную



**Рис. 1.** Организация апикальной меристемы корня *Arabidopsis thaliana*.

Вертикальной линией на схеме отмечена центральная ось корня.

стенку (Heo et al., 2014). Особенность цитологического строения ситовидных элементов заключается в том, что в них отсутствует большинство органоидов (в том числе ядро, вакуоль, шероховатый эндоплазматический ретикулум, комплекс Гольджи, цитоскелет, рибосомы), наличие которых могло бы препятствовать транспорту веществ. Поддержание жизнеспособности ситовидных элементов обеспечивают расположенные в непосредственном контакте с ними клетки-спутницы – паренхимные клетки с крупными ядрами и митохондриями. Механические элементы флоэмы – волокна и склереиды – отличаются друг от друга формой клеток: первые сильно вытянуты в длину и заострены на концах, вторые могут быть лишь слегка удлиненными.

Организация сосудистой системы различается в разных органах, у разных видов и на разных этапах развития растений (Scarpella, Meijer, 2004; Lucas et al., 2013; Furuta et al., 2014). Тем не менее механизмы, определяющие ее развитие, достаточно консервативны (Li et al., 2010; Seo et al., 2020). Так как растительные клетки не способны к миграции, формирование архитектуры органов и тканей в процессе морфогенеза происходит за счет регуляции последовательности и ориентации клеточных делений. Анатомически процесс развития сосудистой системы изучен достаточно хорошо и подробно описан (Scheres et al., 1994; Evert, Eichhorn, 2006; Miyashima et al., 2013; Furuta et al., 2014; De Rybel et al., 2014b, 2016), однако контролирующие его молекулярно-генетические механизмы известны гораздо меньше. Наши представления о них основаны преимущественно на результатах исследований модельного растения *Arabidopsis thaliana*.

В следующих разделах мы приведем краткое описание гистогенеза проводящей ткани у *A. thaliana*, а также реконструируем соответствующую гистогенезу последовательность регуляторных событий. Рассмотрим контроль развития сосудистой системы корня в зародыше и проростке, т. е. на самых ранних этапах ее формирования. Описание механизмов, регулирующих последующие этапы, можно найти в обзорах (Agustí, Blázquez, 2020; Seo et al., 2020).

### Гистогенез проводящих тканей корня

Онтогенетически различают первичные (продуцируемые *первичной меристемой*) и вторичные (продуцируемые *вторичной меристемой*) проводящие ткани.

### Развитие первичных проводящих тканей корня

У *A. thaliana* в раннем эмбриогенезе на стадии глобулы происходит спецификация четырех *проваскулярных инициалей*, которые в результате роста и скоординированных во времени ориентированных делений дают начало *проваскулярной меристеме* зародышевого корня и гипокотилля (рис. 2) (Scheres et al., 1994; Evert, Eichhorn, 2006; Miyashima et al., 2013; Furuta et al., 2014; De Rybel et al., 2014b, 2016). Ее клетки не дифференцированы, однако клеточная судьба некоторых из них уже определена – после прорастания зародыша они дадут начало либо клеткам ксилемы, либо клеткам флоэмы. Расположение этих преддетерминированных клеток в проваскулярной меристеме совпадает с бисимметричной (т. е. обладающей двумя плоскостями симметрии) *диархной* организацией

### Словарь терминов

**Амфирибральный проводящий пучок** – проводящий пучок, в котором флоэма окружает ксилему.

**Антиклинальное клеточное деление** – деление клеток в антиклинальной плоскости. Оно приводит к росту в длину.

**Антиклинальный** – располагающийся в плоскости, перпендикулярной поверхности ткани или органа. Говоря об антиклинальных клеточных стенках или делениях, мы будем подразумевать антиклинальную плоскость, перпендикулярную центральной оси органа.

**Апикальная меристема корня** – первичная меристема корня, которая дает начало всем клеткам корня в процессе его роста в постэмбриональный период.

**Асимметричное клеточное деление** – деление клетки, в результате которого образуются две дочерние клетки с различной клеточной судьбой.

**Вторичная меристема** – меристема, которая закладывается в постэмбриональный период.

**Гипофиза** – верхняя клетка суспензора, которая приобретает свою идентичность на стадии 16–32 клеток; дает начало покоящемуся центру (организующему центру апикальной меристемы корня) и корневому чехлику.

**Диархный проводящий пучок** – проводящий пучок, в котором флоэма и ксилема расположены на разных радиусах; при этом выделяется два луча ксилемы.

**Ксилемная пластинка** – расположенные слоем в центральной плоскости вдоль оси корня клетки первичной ксилемы (или их преддетерминированные предшественники).

**Первичная кора** – периферическая, лежащая снаружи от стелы часть молодого корня.

**Первичная меристема** – меристема, формирующаяся в эмбриогенезе.

**Периклинальное клеточное деление** – деление клеток в периклинальной плоскости. Приводит к увеличению количества клеточных слоев в радиальном направлении.

**Периклинальный** – располагающийся в плоскости, параллельной поверхности ткани или органа.

**Перицикл** – слой паренхимных клеток, окружающих проводящие ткани, наружный слой стелы.

**Проваскулярная меристема корня/гипокотилля** – первичная меристема зародышевого корня и гипокотилля, из которой дифференцируется первичная сосудистая система этих органов после прорастания зародыша.

**Проваскулярные инициали** – четыре клетки проэмбриона, возникающие на стадии ранней глобулы, из которых происходит вся проваскулярная меристема корня/гипокотилля, и только она.

**Прокамбий** – недетерминированные клетки первичной сосудообразующей меристемы, которые в корне *Arabidopsis thaliana* располагаются между ксилемной пластинкой и флоэмными полюсами.

**Сосудистый камбий** – вторичная сосудообразующая меристема, обеспечивающая рост корня в толщину.

**Стела (центральный цилиндр)** – расположенный в центре осевого органа комплекс первичных проводящих тканей, окруженных перициклом.

**Суспензор** – структура у основания зародыша, соединяющая его с эндоспермом. Состоит из потомков базальной клетки двуклеточного проэмбриона.

**Эндодерма** – самый внутренний слой клеток первичной коры, окружающий стелу.

сосудистой системы кончика корня в постэмбриональный период развития: на поперечном срезе по диаметру располагается один слой клеток-предшественников ксилемы, к которому с обеих сторон прилегают клетки **прокамбия**, отделяя будущую ксилему от двух тяжей клеток-предшественников флоэмы, расположенных в перпендикулярной плоскости (Dolan et al., 1993) (рис. 2 и 3, а). Эта структура окружена клетками **перицикла**, которые также происходят от проваскулярных инициалей. Вместе они составляют **центральный цилиндр (стелу)** (см. рис. 1). Стоит отметить, что номенклатура стадий развития проводящей ткани корня достаточно размыта (Furuta et al., 2014), в частности, во многих источниках термин «прокамбий» используют для обозначения не только недетерминированных клеток первичной проводящей ткани проростка, но и всей проваскулярной меристемы зародыша (например, (Busse, Evert, 1999)).

Вскоре после прорастания дифференцировка проводящих элементов начинается в стеле гипокотыля и жилках семядолей (проваскулярная меристема последних происходит от апикальной меристемы побега (Miyashima et al., 2013)). От гипокотыля этот процесс распространяется вверх и вниз по эпикотылю и корню соответственно (Busse, Evert, 1999; De Rybel et al., 2014b; Furuta et al., 2014). У *A. thaliana* первыми дифференцируются прилегающие к перициклу ситовидные элементы протофлоэмы. Поскольку окружающие их клетки продолжают активно расти в длину, клетки протофлоэмы вскоре погибают, и их функционально замещают дифференцирующиеся позднее ситовидные элементы метафлоэмы, которые располагаются ближе к центру стелы (Graeff, Hardtke, 2021; Truernit, 2022). Далее формируются проводящие элементы протоксилемы, расположенные на полюсах **ксилемной пластинки** и имеющие кольчатые или спиральные утолщения вторичных клеточных стенок. Последними дифференцируются клетки метаксилемы с точечно-порвыми или ретикулярными отложениями лигнина, занимающие центральное положение в ксилемной пластинке (Růžička et al., 2015).

В процессе роста корня в длину новые клетки продуцируются за счет **антиклинальных делений** клеток **апикальной меристемы**, расположенной в кончике корня (Desvoyes et al., 2021). Апикальная меристема корня *A. thaliana* закрытого типа, т. е. разные стволовые клетки (инициали), имея неограниченный потенциал деления, продуцируют не любой, а строго ограниченный тип клеток, и для каждой дифференцированной клетки легко проследить, от какой стволовой клетки она произошла (см. рис. 1). Среди инициалей стелы можно выделить те, которые дают начало: 1) протоксилеме; 2) метаксилеме; 3) клеткам прокамбия и ситовидным элементам прото- и метафлоэмы (в этом случае три типа клеток происходят от одной стволовой клетки в результате серии антиклинальных и **периклинальных делений**); 4) только клеткам прокамбия; 5) перициклу (см. рис. 3, б) (Mähönen et al., 2000; Rodriguez-Villalon et al., 2015; Truernit, 2022). Взаимное расположение инициалей соответствует диархной организации сосудистой системы корня (см. рис. 3). Таким образом, клеточная идентичность, установленная в проваскулярной меристеме зародыша, поддерживается

в апикальной меристеме корня. Следует отметить, что в сосудистой системе помимо прото- и метаксилемы, прото- и метафлоэмы и прокамбия выделяют также клетки-спутницы. Некоторые авторы предлагают более строго определять их как клетки, подобные клеткам-спутницам (Truernit, 2022). Эти клетки прилегают к ситовидным элементам прото- и метафлоэмы и обладают рядом морфологических и физиологических характеристик клеток-спутниц (Stadler et al., 2005; Ross-Elliott et al., 2017; Smetana et al., 2019; Graeff, Hardtke, 2021), но, в отличие от последних, в нише стволовых клеток они не имеют общей инициали с элементами прото- и метафлоэмы (Mähönen et al., 2000). Клетки, подобные клеткам-спутницам, дифференцируются, когда начинают функционировать ситовидные элементы протофлоэмы (Graeff, Hardtke, 2021). У *A. thaliana* паренхима ксилемы и флоэмы, волокна, а также истинные клетки-спутницы дифференцируются только во время вторичного роста (Růžička et al., 2015; Truernit, 2022).

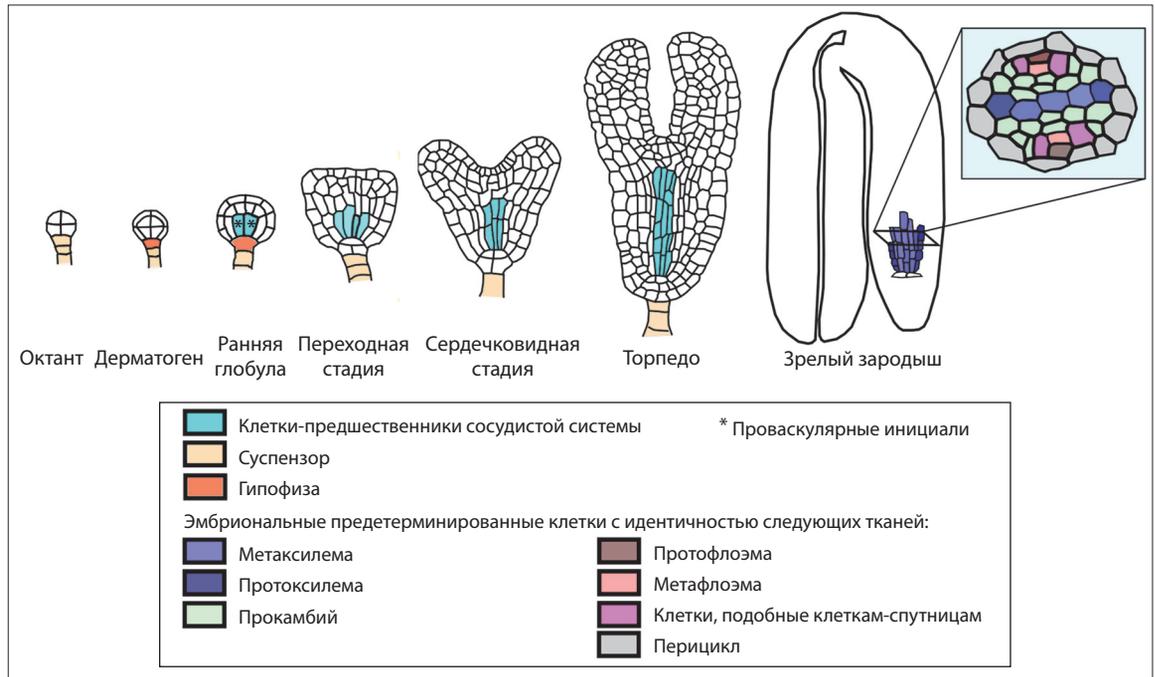
### Формирование камбия

В первичной сосудистой системе *A. thaliana* периклинальные деления клеток прокамбия немногочисленны, однако по мере развития корня после формирования первичной сосудистой сети эти клетки начинают активно делиться в периклинальной плоскости. Периклинальные деления происходят также в клетках перицикла, прилегающих к ксилемной пластинке. В результате формируется замкнутое кольцо клеток, окружающих ксилему, которое дает начало **сосудистому камбию** (см. рис. 3, в) (Baum et al., 2002; Nieminen et al., 2015; Růžička et al., 2015; Smetana et al., 2019). Только те клетки прокамбия и перицикла, которые имеют непосредственный контакт с первичными сосудами ксилемы, дают начало сосудистому камбию (т. е. обладают свойствами стволовых) (Smetana et al., 2019). Потомки других пролиферирующих клеток прокамбия дифференцируются в клетки флоэмы.

Таким образом, в процессе роста корня диархная организация первичной сосудистой системы заменяется **амфикириальной** организацией вторичной сосудистой системы, когда ксилема окружена флоэмой, а между ними располагается слой камбия (см. рис. 3, в). Путем **асимметричных делений** каждая стволовая клетка камбия способна продуцировать наружу клетки флоэмы, а внутрь – клетки ксилемы, обеспечивая рост корня в толщину (Smetana et al., 2019). У некоторых видов, например у большинства однодольных, камбий не образуется и вторичный рост отсутствует. В этом случае все клетки прокамбия в конечном итоге дифференцируются.

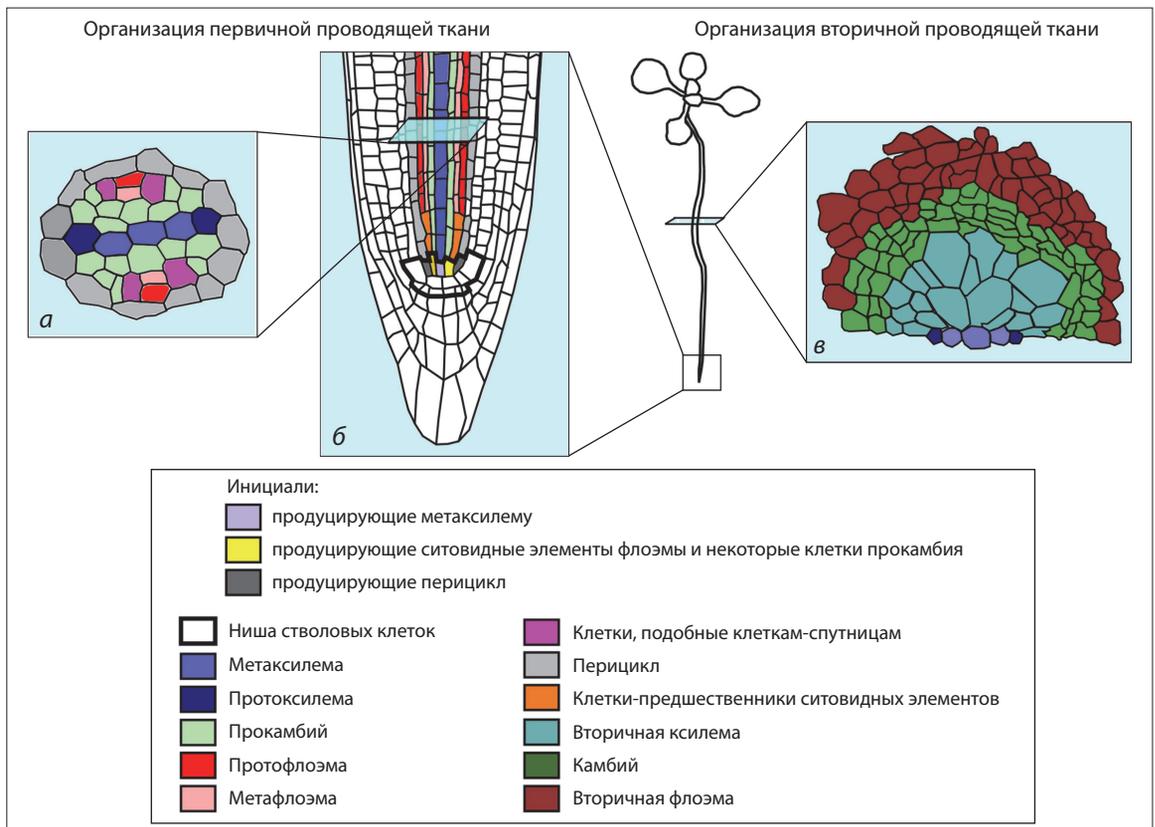
### Установление полярности зародыша и преддетерминация проваскулярных инициалей

Развитие многоклеточного организма сопровождается постепенным ростом ограничения клеточных потенциалов. Первый этап этого процесса – преддетерминация, или спецификация, т. е. определение судьбы тотипотентной клетки, установление, предшественником клеток какого типа она станет. При этом клетка остается недифференцированной, и ее спецификация при определенных условиях обратима. Детерминация клеточной идентичности задействи-



**Рис. 2.** Развитие проваскулярной меристемы *A. thaliana* в эмбриогенезе.

Зрелый зародыш содержит преддетерминированные, но не дифференцированные клетки-предшественники элементов сосудистой системы.



**Рис. 3.** Организация первичной и вторичной проводящей ткани в корне *A. thaliana* в постэмбриональный период.

Стела кончика корня (*a*, *б*) имеет диархную организацию и состоит из прокамбия, первичной флоэмы и первичной ксилемы, окруженных перициклом. Первичная флоэма состоит из прото- и метафлоэмы, а также клеток, подобных клеткам-спутницам. Первичная ксилема состоит из прото- и метаксилемы. В нише стволовых клеток на продольном срезе (*б*) видны две инициали, продуцирующие прокамбий/прото-/метафлоэму, две инициали, продуцирующие перицикл, и одна инициаль, продуцирующая метаксилему. При переходе корня ко вторичному росту (*в*) камбий продуцирует наружу клетки флоэмы, а внутрь – клетки ксилемы, в результате диархная организация проводящего пучка заменяется амфикирибальной.

ет локальное накопление сигнальных молекул, которые активируют или подавляют работу генных сетей, присущих конкретным клеточным типам. В данном процессе важную роль, как правило, играют мобильные факторы, способные перемещаться между клетками и формировать градиенты (Seo et al., 2020).

Спецификации проваскулярных инициалей на ранней глобулярной стадии эмбриогенеза предшествует серия клеточных делений и установление полярности зародыша (Lau et al., 2012; De Rybel et al., 2014b). Корректное прохождение этих процессов необходимо, чтобы развитие проводящей ткани началось с правильного количества клеток, занимающих правильную позицию. Ключевым регулятором эмбриогенеза служит фитогормон ауксин, неоднородное распределение которого предоставляет позиционную информацию, направляющую развитие зародыша (Weijers, Jürgens, 2005; Smit, Weijers, 2015; Mironova et al., 2017). Основным эффектором ауксина в эмбриогенезе является транскрипционный фактор (ТФ) AUXIN RESPONSE FACTOR 5 (ARF5)/MONOPTEROS (MP) (Smit, Weijers, 2015; Verma et al., 2021). Считается, что формирование паттерна распределения сигнала ауксина обеспечивается главным образом за счет обратных связей в регуляции биосинтеза фитогормона, его полярного межклеточного транспорта и сигнального пути (Sauer et al., 2006; Möller, Weijers, 2009; Lau et al., 2011; Robert et al., 2015). В результате на ранних стадиях эмбриогенеза ауксин накапливается в апикальных клетках, определяя таким образом полярность зародыша (Wabnik et al., 2013). Начиная с ранней глобулярной стадии (32 клетки) его максимум смещается в верхние клетки *суспензора*, включая *гипофизу*, которая в дальнейшем даст начало покоящемуся центру апикальной меристемы корня (Friml et al., 2003; Tanaka et al., 2006).

Хотя анатомически четыре проваскулярные инициали выделяются только на ранней глобулярной стадии (Scheres et al., 1994), клеточная идентичность предшественников проводящей ткани устанавливается в четырех внутренних клетках нижнего слоя проэмбриона уже на стадии дерматогена (рис. 4, а) (Smit et al., 2020). В результате периклиналиного деления при переходе на стадию 32 клеток они продуцируют наружу клетки-предшественники *первичной коры*, которые теряют сосудистую идентичность их материнских клеток (см. рис. 4, а, б) (Palovaara et al., 2017; Smit et al., 2020). Необходимым условием спецификации проваскулярных инициалей является ARF5/MP-зависимая активация сигнального пути ауксина, однако этого условия недостаточно (Möller et al., 2017; Smit et al., 2020). Конкретные участники, содействующие ауксину, достоверно неизвестны, хотя предполагается, что эту роль выполняет не один ключевой регулятор, а многокомпонентная регуляторная сеть, одним из участников которой может быть ТФ G-BOX BINDING FACTOR 2 (GBF2) (Smit et al., 2020) (см. рис. 4, а). Считается, что GBF2 модулирует связывание ARF5/MP с промотором генов-мишеней. Стоит отметить, что состояние, в котором единообразно специфицированы предшественники сосудистой системы, вероятнее всего, кратковременное, без устойчивой единообразной клеточной идентичности.

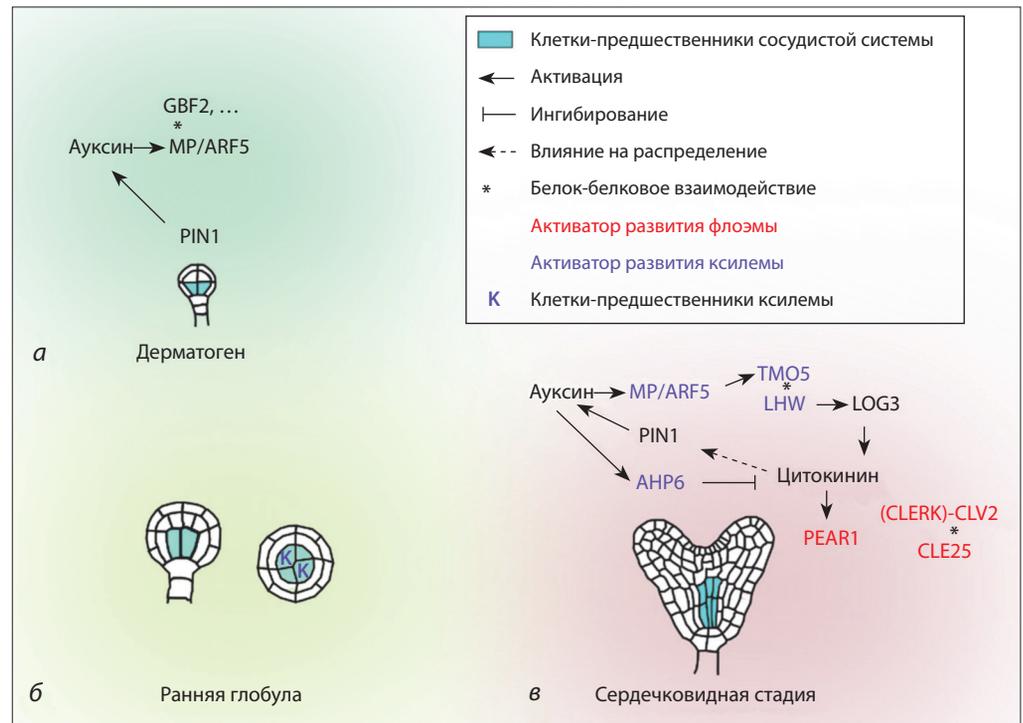
## Предетерминация клеток сосудистой системы в проваскулярной меристеме

По мере контролируемых ориентированных делений проваскулярных инициалей и их потомков происходит разметка сосудистой системы гипокотыля и корня путем спецификации конкретных клеточных типов. На этом этапе важным аспектом является установление границ клеточных доменов, имеющих разную структурно-функциональную принадлежность. К концу эмбриогенеза в проваскулярной меристеме зародыша уже определена клеточная идентичность всех элементов сосудистой системы: протоксилемы, метаксилемы, протофлоэмы, метафлоэмы, клеток, подобных клеткам-спутницам, и прокампбия, о чем в совокупности свидетельствуют данные по морфологии клеток и экспрессии генов-маркеров (см. рис. 2) (Bonke et al., 2003; Bauby et al., 2007).

Считается, что уже на ранней глобулярной стадии эмбриогенеза бисимметрию будущего корня *A. thaliana* предопределяет наличие расширенного контакта между двумя расположенными по диагонали друг относительно друга проваскулярными инициалами (см. рис. 4, б). Такой контакт возникает, вероятно, в результате неточного совпадения плоскостей деления в проэмбрионе уже на стадии четырех клеток и в дальнейшем важен для формирования ксилемной пластинки (De Rybel et al., 2014a). Начиная с ранней сердечковидной стадии эмбриогенеза в такие контактирующие проваскулярные клетки активно транспортируется ауксин из расположенных над ними зачатков семядолей, тогда как в других клетках уровень гормона остается низким (Bishopp et al., 2011a; Help et al., 2011; De Rybel et al., 2014a). Локальное повышение концентрации ауксина необходимо для спецификации клеток-предшественников ксилемы (Bishopp et al., 2011a).

В то же время эти клетки благодаря высокому содержанию ауксина выступают в качестве организующего центра проваскулярной меристемы, который скоординированно регулирует ее рост за счет периклиналиных делений и установления бисимметричной организации (De Rybel et al., 2014a). Ауксин индуцирует ARF5/MP-зависимую экспрессию ТФ TARGET OF MONOPTEROS 5 (TMO5) и TMO5-LIKE1 (T5L1) (Schlereth et al., 2010; De Rybel et al., 2013, 2014b), которые, формируя гетеродимеры с ауксин-независимым ТФ LONESOME HIGHWAY (LHW) (De Rybel et al., 2013), активируют экспрессию генов биосинтеза цитокинина *LONELY GUY3 (LOG3)* и *LOG4* (Kuroha et al., 2009; De Rybel et al., 2014a) (см. рис. 4, в). Одновременно ауксин блокирует передачу цитокининового сигнала, повышая экспрессию гена *ARABIDOPSIS HISTIDINE PHOSPHOTRANSFER PROTEIN 6 (AHP6)*, кодирующего ингибитор сигнального пути цитокинина (Mähönen et al., 2006; Bishopp et al., 2011a). Таким образом, в клетках-предшественниках ксилемы формируется локальный источник цитокинина на фоне блокировки сигнального пути этого фитогормона.

Повышенный уровень цитокинина, с одной стороны, ограничивает утечку ауксина из клеток-предшественников ксилемы, контролируя локализацию транспортера ауксина PIN-FORMED 1 (PIN1) на клеточной мембране (Marhavý et al., 2011; De Rybel et al., 2014a). С другой стороны, цитокинин посредством диффузии мигрирует



**Рис. 4.** Генетическая регуляция развития проваскулярной меристемы в эмбриогенезе.

*a* – преддетерминация проваскулярных инициалей. Клеточная идентичность предшественников проводящей ткани устанавливается в четырех внутренних клетках нижнего слоя проэмбриона на стадии дерматогена, хотя анатомически четыре проваскулярные инициали выделяются только на ранней глобулярной стадии; *б* – преддетерминация клеток-предшественников ксилемы на ранней глобулярной стадии; *в* – формирование бисимметричного паттерна и преддетерминация клеток-предшественников ксилемы и флоэмы начиная с сердцевидной стадии.

в соседние клетки по градиенту концентрации. В этих клетках активация сигнального каскада цитокинина в отсутствие ингибитора (Cheng, Kieber, 2014) стимулирует периклинальные деления (Smit, Weijers, 2015). В то же время активация сигнального пути цитокинина подавляет спецификацию клеток по ксилемному пути развития (Mähönen et al., 2006). Этот механизм обеспечивает радиальный рост проваскулярной меристемы, который сопровождается пространственным разделением доменов повышенного сигнала ауксина (клетки приобретают идентичность ксилемы) и цитокинина (плюрипотентные клетки прокамбия). Его достаточность для самоорганизации бисимметричного паттерна в проваскулярной меристеме была подтверждена с помощью математической модели (De Rybel et al., 2014a).

В клетках-предшественниках проваскулярной меристемы в раннем эмбриогенезе начинают экспрессироваться гены, кодирующие пептидный гормон CLAVATA 3 (CLV3)/EMBRYO SURROUNDING REGION 25 (CLE25) и мобильные ТФ семейства DNA BINDING WITH ONE FINGER (DOF), объединенные в группу PHLOEM EARLY DOF (PEAR), которые в постэмбриональный период маркируют клетки-предшественники ситовидных элементов (Miyashima et al., 2019; Ren et al., 2019). CLE25 экспрессируется начиная со стадии 64-клеточного зародыша (Ren et al., 2019). Независимая от цитокинина активация PEAR1 детектируется уже на стадии 16 клеток, а с ранней сердцевидной стадии экспрессию этого гена активирует

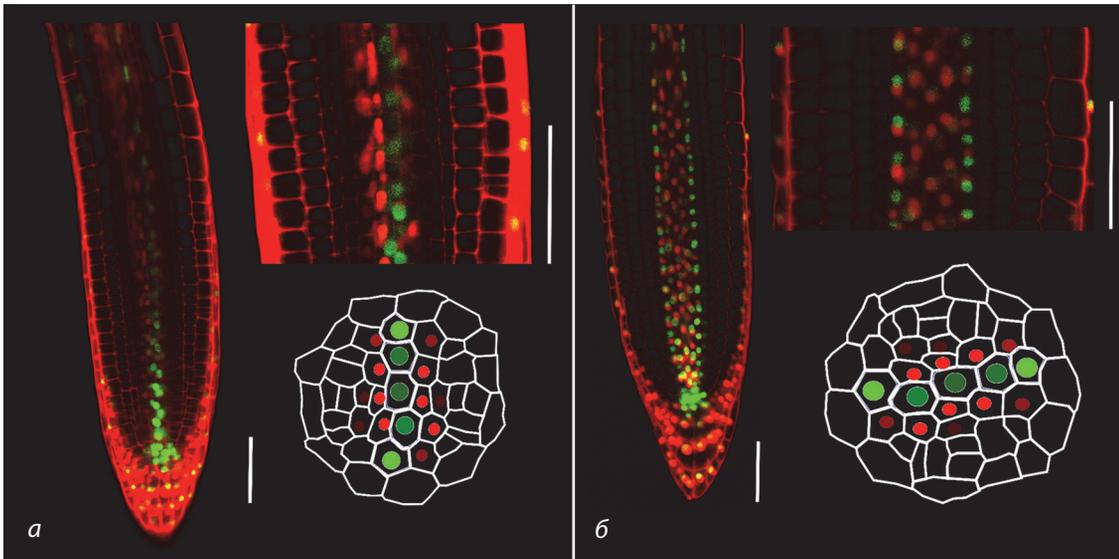
и поддерживает цитокинин (Miyashima et al., 2019). Предполагается, что пептид CLE25, связываясь с рецепторным комплексом CLE-RESISTANT RECEPTOR KINASE (CLERK)-CLV2, и параллельно ТФ PEAR1 вносят вклад в раннюю спецификацию клеток-предшественников флоэмы. Однако, в отличие от клеток ксилемы, механизм, инициирующий развитие флоэмы в эмбриогенезе, пока неизвестен.

### Поддержание клеточной идентичности предшественников ксилемы и флоэмы в апикальной меристеме корня

#### Бисимметричный паттерн в стеле

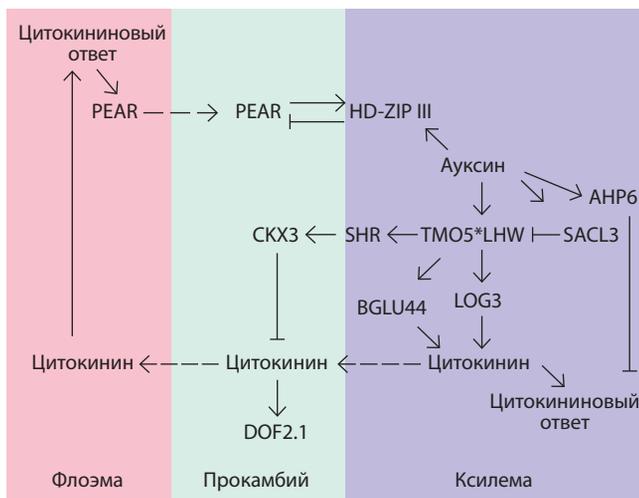
В постэмбриональный период в предшественниках клеток стелы у *A. thaliana* в апикальной меристеме поддерживается бисимметричный паттерн, установленный в эмбриогенезе. Некоторые механизмы, регулирующие клеточную динамику и преддетерминацию элементов сосудистой системы в проваскулярной меристеме, продолжают функционировать и после прорастания. Однако нельзя с полной уверенностью утверждать, что эти механизмы абсолютно идентичны.

В апикальной меристеме клетки-предшественники ксилемы с высоким содержанием ауксина сохраняют функцию организующего центра, осуществляя TMO5/LHW-опосредованную регуляцию уровня цитокинина в клетках прокамбия (рис. 5) (Ohashi-Ito, Bergmann, 2007;



**Рис. 5.** Бисимметричный паттерн распределения ауксина и цитокинина в стеле кончика корня *A. thaliana*. Ксилемная пластинка располагается перпендикулярно (а) или параллельно (б) плоскости оптического среза.

Микроизображения для репортерной линии *TCSn::ntdTomato-DR5revV2::n3GFP* (Smet et al., 2019) получены нами с помощью конфокального микроскопа. Клеточные стенки окрашены пропидий йодидом. Ядерные сигналы GFP (зеленый) и Tomato (красный) отражают активность сигнальных путей ауксина и цитокинина соответственно. Ауксиновый ответ наблюдается в предшествениках ксилемы с максимумом в предшествениках протоксилемы, цитокининовый ответ – в клетках прокамбия, контактирующих с ксилемой, маркируя таким образом морфофункциональные домены стелы кончика корня. Шкала 50 мкм.



**Рис. 6.** Механизм поддержания бисимметричного паттерна распределения ауксина и цитокинина в стеле кончика корня *A. thaliana* в постэмбриональный период.

Звездочка обозначает физическое взаимодействие между белками с формированием димера, штриховая стрелка – перемещение мобильных регуляторов.

Bishopp et al., 2011a; De Rybel et al., 2013; Ohashi-Ito et al., 2013, 2014; Vera-Sirera et al., 2015; Yang et al., 2021). Высокое содержание активного цитокинина в клетках ксилемы поддерживается путем TMO5/LHW-зависимой активации не только генов биосинтеза цитокинина *LOG3* и *LOG4*, но и гена *BGLU44*, кодирующего фермент β-глюкозидазу (рис. 6). Цитокининовый ответ в ксилеме блокируется ауксином посредством индукции экспрессии гена *AHP6* (Bishopp et al., 2011a), а также ограничения активности

TMO5/LHW путем активации регуляторного модуля *ACAULIS 5 (ACL5)–SUPPRESSOR OF ACAULIS5 LIKE3 (SACL3)*, который блокирует формирование гетеродимера TMO5/LHW в результате конкуренции ТФ SACL3 за связывание с LHW (Katayama et al., 2015; Cai et al., 2016) (см. рис. 6). В прилегающих к ксилеме клетках прокамбия уровень цитокинина, диффундирующего из ксилемы, ограничивается TMO5/LHW-зависимой активацией гена цитокинин-оксидазы *CYTOKININ OXIDASE 3 (CKX3)*. Активация этого гена опосредуется мобильным ТФ *SHORT ROOT (SHR)* – продуктом гена-мишени гетеродимера TMO5/LHW. Сочетанное действие разнонаправленных регуляторных модулей обеспечивает устойчивость паттерна к кратковременным колебаниям концентраций ауксина в клетках ксилемы, сохраняя его чувствительность к более медленным/стабильным изменениям (Yang et al., 2021). Интересно, что ген *SHR* важен не только для организации радиальной симметрии корня, но и для функционирования покоящегося центра (Творогова и др., 2012).

TMO5/LHW-индуцированный цитокинин активирует транскрипцию ТФ *DOF2.1* в клетках прокамбия в районе ксилемного полюса, контролируя таким образом деление этих клеток (см. рис. 6) (Smet et al., 2019). Помимо клеток ксилемы, источником цитокинина в апикальной меристеме корня может служить дифференцированная флоэма, по которой транспортируется фитогормон (Bishopp et al., 2011b). Однако, согласно результатам математического моделирования, флоэмный цитокинин не является основополагающим источником позиционной информации для формирования бисимметричного паттерна (Murgaro et al., 2014). В то же время высокое содержание цитокинина на флоэмных полюсах организует периклинальные деления прокамбия путем активации генов мобильных ТФ семей-

ства DOF, объединяемых в группу PEAR, в том числе PEAR1, PEAR2, TMO6, DOF6 (Miyashima et al., 2019; Smet et al., 2019). Они создают концентрационный градиент и активируют периклиальные деления прокамбиальных клеток, окружающих флоэмный полус. Транскрипционные факторы класса III HOMEODOMAIN LEU-ZIPPER (HD-ZIP III), домен экспрессии которых располагается в центральной части стелы (см. далее), ограничивают активность ТФ группы PEAR (см. рис. 6). PEAR1 активирует транскрипцию генов семейства HD-ZIP III, формируя таким образом петлю отрицательной обратной связи.

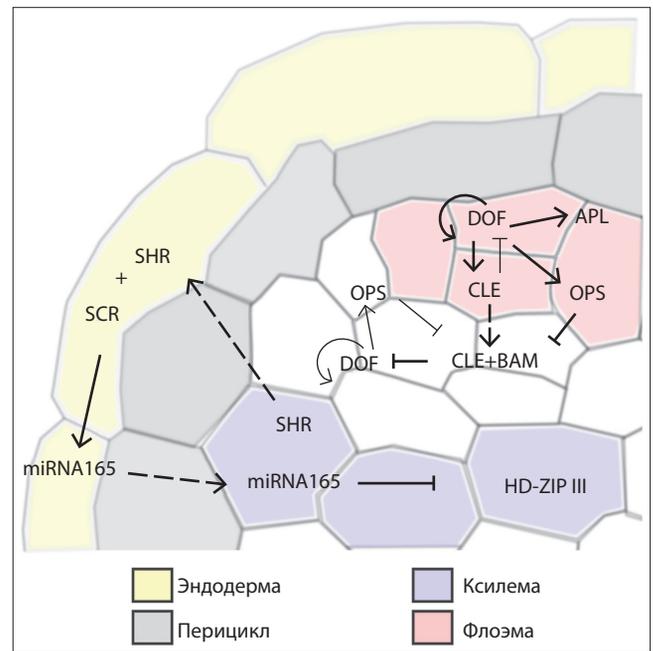
### Преддетерминация прото- и метаксилемы

Как и в эмбриогенезе, ауксин необходим для преддетерминации клеток ксилемы в апикальной меристеме корня. В разграничении прото- и метаксилемы ключевую роль играют мобильные регуляторы SHR и miRNA165/166 (рис. 7). Источником SHR являются клетки ксилемы, откуда ТФ распространяется по направлению к периферии и, достигнув *эндодермы*, активирует ТФ SCARECROW (SCR), в комплексе с которым индуцирует экспрессию miRNA165/166 (Carlsbecker et al., 2010; De Rybel et al., 2016). МикроРНК диффундируют в соседние клетки, формируя градиент концентрации по направлению к центру корня. В стеле miRNA165/166 подавляют экспрессию генов, кодирующих ТФ семейства HD-ZIP III, ограничивая ее центральной областью (см. рис. 7). Таким образом преддетерминируются клетки метаксилемы. Работает ли данный механизм в эмбриогенезе, неизвестно, однако это не исключено. В частности, ТФ PHABULOSA (PHB) семейства HD-ZIP III экспрессируется в корне зародыша (Grigg et al., 2009).

### Преддетерминация элементов флоэмы

К маркерам флоэмы, которые экспрессируются в клетках-предшественниках и индуцируют развитие ткани, относят целый ряд ТФ семейства DOF (Miyashima et al., 2019; Roszak et al., 2021), супрессоры сигнального пути стригулактонов SUPPRESSOR OF MAX2 1-LIKE 3 (SMLX3), SMLX4 и SMLX5 (Wallner et al., 2017), мембранные белки BREVIS RADIX (BRX), OCTOPUS (OPS), OPS-LIKE 2 (OPL2) (Ruiz Sola et al., 2017), фосфатазу COTYLEDON VASCULAR PATTERN 2 (CVP2), ее гомолог CVP2-LIKE 1 (CVL1) (Rodriguez-Villalon et al., 2015) и ТФ ALTERED PHLOEM DEVELOPMENT (APL) (Bonke et al., 2003). Формирование элементов протофлоэмы контролируется путем смещения баланса в сторону индуцирующих или подавляющих механизмов.

Центральным звеном, связывающим противодействующие регуляторные модули, являются специфичные для флоэмы ТФ семейства DOF (Qian et al., 2022). С одной стороны, они индуцируют экспрессию активаторов развития флоэмы, таких как *APL*, а также собственных генов, формируя петлю положительной обратной связи. С другой стороны, DOF индуцируют экспрессию сигнальных пептидов CLE25, CLE26 и CLE45, которые мигрируют в соседние клетки, где запускают ингибирующий регуляторный модуль (см. рис. 7). Взаимодействуя с рецепторами класса BARELY ANY MERISTEM (BAM) и корцепторами CLAVATA3 INSENSITIVE RECEPTOR



**Рис. 7.** Генетический контур, регулирующий преддетерминацию клеток прото- и метаксилемы и флоэмы в корне *A. thaliana*.

Разделение доменов прото- и метаксилемы определяется концентрационным градиентом ТФ семейства HD-ZIP III. Активируемый ауксином мобильный ТФ SHR диффундирует из ксилемы в эндодерму, где, соединяясь с белком SCR, активирует экспрессию miRNA165. МикроРНК, которые разрушают мРНК ТФ семейства HD-ZIP III, образуют градиент концентрации по направлению к центру. В результате локализация ТФ HD-ZIP III ограничивается центральной областью, преддетерминируя клетки метаксилемы. Преддетерминация флоэмы начинается с цитокинин-активируемой экспрессии ТФ семейства DOF. Они индуцируют сигнальные CLE пептиды, которые мигрируют в соседние клетки, взаимодействуют с рецепторами BAM и индуцируют деградацию DOF. Таким образом создается граница между будущей флоэмой и соседними клетками. Штриховая стрелка обозначает перемещение мобильных регуляторов.

KINASE (CIK), CLE пептиды индуцируют деградацию ТФ семейства DOF, подавляя формирование элементов протофлоэмы. Активность рецепторов CLE пептидов может дополнительно регулироваться, например регуляторами MEMBRANE-ASSOCIATED KINASE REGULATOR 5 (MAKR5) (Kang, Hardtke, 2016) или CORYNE (CRN) (Hazak et al., 2017). Транскрипционные факторы семейства DOF активируют экспрессию генов, кодирующих мембранный белок OPS, который подавляет BAM-CIK модуль (Qian et al., 2022). Правильно позиционированные клетки-предшественники протофлоэмы преодолевают ингибиторный эффект CLE пептидов благодаря аккумуляции ТФ DOF, обусловленной функционированием положительной обратной связи. Такой балансирующий механизм позволяет осуществить репаттернинг флоэмы, если развитие протофлоэмы было нарушено (Gujas et al., 2020). Отметим, что развитие метафлоэмы у *A. thaliana*, вероятно, регулируется другими механизмами и не зависит от развития протофлоэмы (Graeff, Hardtke, 2021).

В процессе формирования флоэмы стволовая клетка флоэмы/прокамбия делится антиклинально, продуцируя дочернюю клетку-предшественник прокамбия и ситовидных элементов, которая делится периклиально с образованием клетки прокамбия и предшественника си-

товидных элементов флоэмы. Последняя в результате еще одного периклиального деления дает предшественники протофлоэмы и метафлоэмы (Rodriguez-Villalon, 2016). Клетки, подобные клеткам-спутницам, тоже формируются в результате асимметричного деления, но происходят от другой стволовой клетки. Эти асимметричные клеточные деления контролируются позиционным сигналом – градиентом белка SHR, миграция которого в эндодерму активирует микроРНК miRNA165/166 и вызывает асимметричные деления, продуцирующие клетки, подобные спутницам, тогда как продвижение SHR во флоэму необходимо для асимметричных делений, ведущих к образованию протоксилемы и метаксилемы (Kim et al., 2020).

## Заключение

Сосудистая система корня *A. thaliana* закладывается на самых ранних стадиях эмбриогенеза. При этом предтерминация проваскулярных инициалей подразумевает лабильную, неустойчивую и обратимую спецификацию, сформированную на основе физического расположения клеток в зародыше под воздействием комплексной регуляторной сети транскрипционных факторов. Интересно, что и маркеры ксилемы (например, TMO5, T5L1), и маркеры флоэмы (например, PEAR1, TMO6, DOF6) совместно экспрессируются в проваскулярных инициалах в раннем эмбриогенезе, но далее в проваскулярной меристеме и в проростке в процессе развития сосудистой системы разделяются по разным пространственным доменам.

Разметка сосудистой системы у *A. thaliana* заканчивается к моменту созревания зародыша. Частично генная сеть, контролирующая этот процесс в эмбриогенезе, продолжает поддерживать структуру сосудистой системы растущего корня в проростке и далее в онтогенезе растения. Эти механизмы ассоциированы с локальным накоплением молекулярных маркеров, устойчиво экспрессирующихся в клетках-предшественниках определенного типа. Тем не менее факторы, работающие и в эмбриогенезе, и в постэмбриональном развитии, могут действовать на данных этапах по-разному.

Несмотря на существенный прогресс в понимании молекулярно-генетических механизмов регуляции развития сосудистой системы, достигнутый в последнее время, остается много вопросов, в частности в связи с существованием параллельных путей регуляции и петель прямой связи. Это является хорошей основой для построения математических моделей, анализ которых поможет пролить свет на взаимоотношения между различными регуляторными контурами и их функциональную значимость.

## Список литературы / References

Творогова В.Е., Осипова М.А., Додуева И.Е., Лутова Л.А. Взаимодействие транскрипционных факторов и фитогормонов в регуляции активности меристем у растений. *Экол. генетика*. 2012; 10(3):28-40.  
[Tvorogova V.E., Osipova M.A., Doduyeva I.E., Lutova L.A. Interaction between transcriptional factors and phytohormones in regulation of plant meristems activity. *Ekologicheskaya Genetika = Ecological Genetics*. 2012;10(3):28-40. (in Russian)]  
Agustí J., Blázquez M.A. Plant vascular development: mechanisms and environmental regulation. *Cell. Mol. Life Sci.* 2020;77(19):3711-3728. DOI 10.1007/s00018-020-03496-w.

Bauby H., Divol F., Truernit E., Grandjean O., Palauqui J.C. Proto-phloem differentiation in early *Arabidopsis thaliana* development. *Plant Cell Physiol.* 2007;48(1):97-109. DOI 10.1093/pcp/pc1045.  
Baum S.F., Dubrovsky J.G., Rost T.L. Apical organization and maturation of the cortex and vascular cylinder in *Arabidopsis thaliana* (Brassicaceae) roots. *Am. J. Bot.* 2002;89(6):908-920. DOI 10.3732/ajb.89.6.908.  
Bishopp A., Help H., El-Showk S., Weijers D., Scheres B., Friml J., Benková E., Mähönen A.P., Helariutta Y. A mutually inhibitory interaction between auxin and cytokinin specifies vascular pattern in roots. *Curr. Biol.* 2011a;21(11):917-926. DOI 10.1016/j.cub.2011.04.017.  
Bishopp A., Lehesranta S., Vátén A., Help H., El-Showk S., Scheres B., Helariutta K., Mähönen A.P., Sakakibara H., Helariutta Y. Phloem-transported cytokinin regulates polar auxin transport and maintains vascular pattern in the root meristem. *Curr. Biol.* 2011b;21(11):927-932. DOI 10.1016/j.cub.2011.04.049.  
Bonke M., Thitamadee S., Mähönen A.P., Hauser M.T., Helariutta Y. APL regulates vascular tissue identity in *Arabidopsis*. *Nature*. 2003;426(6963):181-186. DOI 10.1038/nature02100.  
Busse J.S., Evert R.F. Pattern of differentiation of the first vascular elements in the embryo and seedling of *Arabidopsis thaliana*. *Int. J. Plant Sci.* 1999;160(1):1-13. DOI 10.1086/314098.  
Cai Q., Fukushima H., Yamamoto M., Ishii N., Sakamoto T., Kurata T., Motose H., Takahashi T. The SAC51 family plays a central role in thermospermine responses in *Arabidopsis*. *Plant Cell Physiol.* 2016;57(8):1583-1592. DOI 10.1093/pcp/pcw113.  
Carlsbecker A., Lee J.Y., Roberts C.J., Dettmer J., Lehesranta S., Zhou J., Lindgren O., Moreno-Risueno M.A., Vátén A., Thitamadee S., Campilho A., Sebastian J., Bowman J.L., Helariutta Y., Benfey P.N. Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. *Nature*. 2010;465(7296):316-321. DOI 10.1038/nature08977.  
Cheng C.Y., Kieber J.J. Cytokinin signaling in plants. In: Howell S. (Ed.) *Molecular Biology. The Plant Sciences*. Vol. 2. New York: Springer, 2014;269-289. DOI 10.1007/978-1-4614-7570-5\_14.  
Courtois-Moreau C.L., Pesquet E., Sjödin A., Muñoz L., Bollhöner B., Kaneda M., Samuels L., Jansson S., Tuominen H. A unique program for cell death in xylem fibers of *Populus* stem. *Plant J.* 2009;58(2):260-274. DOI 10.1111/j.1365-3113.2008.03777.x.  
De Rybel B., Adibi M., Breda A.S., Wendrich J.R., Smit M.E., Novák O., Yamaguchi N., Yoshida S., Van Isterdael G., Palovaara J., Nijse B., Boeckschoten M.V., Hooiveld G., Beeckman T., Wagner D., Ljung K., Fleck C., Weijers D. Integration of growth and patterning during vascular tissue formation in *Arabidopsis*. *Science*. 2014a;345(6197):1255215. DOI 10.1126/science.1255215.  
De Rybel B., Breda A.S., Weijers D. Prenatal plumbing – vascular tissue formation in the plant embryo. *Physiol. Plant.* 2014b;151(2):126-133. DOI 10.1111/ppl.12091.  
De Rybel B., Mähönen A.P., Helariutta Y., Weijers D. Plant vascular development: from early specification to differentiation. *Nat. Rev. Mol. Cell Biol.* 2016;17(1):30-40. DOI 10.1038/nrm.2015.6.  
De Rybel B., Möller B., Yoshida S., Grabowicz I., Barbier de Reuille P., Boeren S., Smith R.S., Borst J.W., Weijers D. A bHLH complex controls embryonic vascular tissue establishment and indeterminate growth in *Arabidopsis*. *Dev. Cell.* 2013;24(4):426-437. DOI 10.1016/j.devcel.2012.12.013.  
Desvoyes B., Echevarría C., Gutierrez C. A perspective on cell proliferation kinetics in the root apical meristem. *J. Exp. Bot.* 2021;72(19):6708-6715. DOI 10.1093/jxb/erab303.  
Dolan L., Janmaat K., Willemsen V., Linstead P., Poethig S., Roberts K., Scheres B. Cellular organisation of the *Arabidopsis thaliana* root. *Development*. 1993;119(1):71-84. DOI 10.1242/dev.119.1.71.  
Evert R.F., Eichhorn S.E. *Esau's Plant Anatomy. Meristems, Cells, and Tissues of the Plant Body: Their Structure, Function, and Development*. New Jersey: Wiley, 2006. DOI 10.1002/0470047380.  
Friml J., Vieten A., Sauer M., Weijers D., Schwarz H., Hamann T., Offringa R., Jürgens G. Efflux-dependent auxin gradients establish the

- apical-basal axis of *Arabidopsis*. *Nature*. 2003;426(6963):147-153. DOI 10.1038/nature02085.
- Furuta K.M., Hellmann E., Helariutta Y. Molecular control of cell specification and cell differentiation during procambial development. *Annu. Rev. Plant Biol.* 2014;65:607-638. DOI 10.1146/annurev-arplant-050213-040306.
- Graeff M., Hardtke C.S. Metaphloem development in the *Arabidopsis* root tip. *Development*. 2021;148(18):dev199766. DOI 10.1242/dev.199766.
- Grigg S.P., Galinha C., Kornet N., Canales C., Scheres B., Tsiantis M. Repression of apical homeobox genes is required for embryonic root development in *Arabidopsis*. *Curr. Biol.* 2009;19(17):1485-1490. DOI 10.1016/j.cub.2009.06.070.
- Gujas B., Kastanaki E., Sturchler A., Cruz T.M.D., Ruiz-Sola M.A., Dreos R., Eicke S., Truemit E., Rodriguez-Villalon A. A reservoir of pluripotent phloem cells safeguards the linear developmental trajectory of protophloem sieve elements. *Curr. Biol.* 2020;30(5):755-766. DOI 10.1016/j.cub.2019.12.043.
- Hazak O., Brandt B., Cattaneo P., Santiago J., Rodriguez-Villalon A., Hothorn M., Hardtke C.S. Perception of root-active CLE peptides requires CORYNE function in the phloem vasculature. *EMBO Rep.* 2017;18(8):1367-1381. DOI 10.15252/embr.201643535.
- Help H., Mähönen A.P., Helariutta Y., Bishopp A. Bisymmetry in the embryonic root is dependent on cotyledon number and position. *Plant Signal. Behav.* 2011;6(11):1837-1840. DOI 10.4161/psb.6.11.17600.
- Heo J.O., Roszak P., Furuta K.M., Helariutta Y. Phloem development: current knowledge and future perspectives. *Am. J. Bot.* 2014;101(9):1393-1402. DOI 10.3732/ajb.1400197.
- Kang Y.H., Hardtke C.S. *Arabidopsis* MAKR5 is a positive effector of BAM3-dependent CLE45 signaling. *EMBO Rep.* 2016;17(8):1145-1154. DOI 10.15252/embr.201642450.
- Katayama H., Iwamoto K., Kariya Y., Asakawa T., Kan T., Fukuda H., Ohashi-Ito K. A negative feedback loop controlling bHLH complexes is involved in vascular cell division and differentiation in the root apical meristem. *Curr. Biol.* 2015;25(23):3144-3150. DOI 10.1016/j.cub.2015.10.051.
- Kim H., Zhou J., Kumar D., Jang G., Ryu K.H., Sebastian J., Miyashima S., Helariutta Y., Lee J.Y. SHORTROOT-mediated intercellular signals coordinate phloem development in *Arabidopsis* roots. *Plant Cell*. 2020;32(5):1519-1535. DOI 10.1105/tpc.19.00455.
- Kuroha T., Tokunaga H., Kojima M., Ueda N., Ishida T., Nagawa S., Fukuda H., Sugimoto K., Sakakibara H. Functional analyses of *LONELY GUY* cytokinin-activating enzymes reveal the importance of the direct activation pathway in *Arabidopsis*. *Plant Cell*. 2009;21(10):3152-3169. DOI 10.1105/tpc.109.068676.
- Lau S., De Smet I., Kolb M., Meinhardt H., Jürgens G. Auxin triggers a genetic switch. *Nat. Cell Biol.* 2011;13(5):611-615. DOI 10.1038/ncb2212.
- Lau S., Slane D., Herud O., Kong J., Jürgens G. Early embryogenesis in flowering plants: setting up the basic body pattern. *Annu. Rev. Plant Biol.* 2012;63:483-506. DOI 10.1146/annurev-arplant-042811-105507.
- Li X., Wu H.X., Southerton S.G. Comparative genomics reveals conservative evolution of the xylem transcriptome in vascular plants. *BMC Evol. Biol.* 2010;10:190. DOI 10.1186/1471-2148-10-190.
- Lucas W.J., Groover A., Lichtenberger R., Furuta K., Yadav S.R., Helariutta Y., He X.Q., Fukuda H., Kang J., Brady S.M., Patrick J.W., Sperry J., Yoshida A., López-Millán A.F., Grusak M.A., Kachroo P. The plant vascular system: evolution, development and functions. *J. Integr. Plant Biol.* 2013;55(4):294-388. DOI 10.1111/jipb.12041.
- Mähönen A.P., Bishopp A., Higuchi M., Nieminen K.M., Kinoshita K., Törmäkangas K., Ikeda Y., Oka A., Kakimoto T., Helariutta Y. Cytokinin signaling and its inhibitor AHP6 regulate cell fate during vascular development. *Science*. 2006;311(5757):94-98. DOI 10.1126/science.1118875.
- Mähönen A.P., Bonke M., Kauppinen L., Riikonen M., Benfey P.N., Helariutta Y. A novel two-component hybrid molecule regulates vascular morphogenesis of the *Arabidopsis* root. *Genes Dev.* 2000;14(23):2938-2943. DOI 10.1101/gad.189200.
- Marhavý P., Bielach A., Abas L., Abuzeineh A., Duclercq J., Tanaka H., Pařezová M., Petrášek J., Friml J., Kleine-Vehn J., Benková E. Cytokinin modulates endocytic trafficking of PIN1 auxin efflux carrier to control plant organogenesis. *Dev. Cell*. 2011;21(4):796-804. DOI 10.1016/j.devcel.2011.08.014.
- Ménard D., Pesquet E. Cellular interactions during tracheary elements formation and function. *Curr. Opin. Plant Biol.* 2015;23:109-115. DOI 10.1016/j.pbi.2014.12.001.
- Mironova V., Teale W., Shahriari M., Dawson J., Palme K. The systems biology of auxin in developing embryos. *Trends Plant Sci.* 2017;22(3):225-235. DOI 10.1016/j.tplants.2016.11.010.
- Miyashima S., Roszak P., Sevilem I., Toyokura K., Blob B., Heo J.O., Mellor N., Help-Rinta-Rahko H., Otero S., Smet W., Boekschooten M., Hooiveld G., Hashimoto K., Smetana O., Siligato R., Wallner E.S., Mähönen A.P., Kondo Y., Melnyk C.W., Greb T., Nakajima K., Sozzani R., Bishopp A., De Rybel B., Helariutta Y. Mobile PEAR transcription factors integrate positional cues to prime cambial growth. *Nature*. 2019;565(7740):490-494. DOI 10.1038/s41586-018-0839-y.
- Miyashima S., Sebastian J., Lee J.Y., Helariutta Y. Stem cell function during plant vascular development. *EMBO J.* 2013;32(2):178-193. DOI 10.1038/emboj.2012.301.
- Möller B.K., ten Hove C.A., Xiang D., Williams N., López L.G., Yoshida S., Smit M., Datla R., Weijers D. Auxin response cell-autonomously controls ground tissue initiation in the early *Arabidopsis* embryo. *Proc. Natl. Acad. Sci. USA*. 2017;114(12):2533-2539. DOI 10.1073/pnas.1616493114.
- Möller B., Weijers D. Auxin control of embryo patterning. *Cold Spring Harb. Perspect. Biol.* 2009;1(5):a001545. DOI 10.1101/cshperspect.a001545.
- Muraro D., Mellor N., Pound M.P., Help H., Lucas M., Chopard J., Byrne H.M., Godin C., Hodgman T.C., King J.R., Pridmore T.P., Helariutta Y., Bennett M.J., Bishopp A. Integration of hormonal signaling networks and mobile microRNAs is required for vascular patterning in *Arabidopsis* roots. *Proc. Natl. Acad. Sci. USA*. 2014;111(2):857-862. DOI 10.1073/pnas.1221766111.
- Nieminen K., Blomster T., Helariutta Y., Mähönen A.P. Vascular cambium development. *Arabidopsis Book*. 2015;13:e0177. DOI 10.1199/tab.0177.
- Ohashi-Ito K., Bergmann D.C. Regulation of the *Arabidopsis* root vascular initial population by *LONESOME HIGHWAY*. *Development*. 2007;134(16):2959-2968. DOI 10.1242/dev.006296.
- Ohashi-Ito K., Matsukawa M., Fukuda H. An atypical bHLH transcription factor regulates early xylem development downstream of auxin. *Plant Cell Physiol.* 2013;54(3):398-405. DOI 10.1093/pcp/ptc013.
- Ohashi-Ito K., Saegusa M., Iwamoto K., Oda Y., Katayama H., Kojima M., Sakakibara H., Fukuda H. A bHLH complex activates vascular cell division via cytokinin action in root apical meristem. *Curr. Biol.* 2014;24(17):2053-2058. DOI 10.1016/j.cub.2014.07.050.
- Palovaara J., Saiga S., Wendrich J.R., van't Wout Hofland N., van Schayck J.P., Hater F., Mutte S., Sjollem J., Boekschooten M., Hooiveld G.J., Weijers D. Transcriptome dynamics revealed by a gene expression atlas of the early *Arabidopsis* embryo. *Nat. Plants*. 2017;3(11):894-904. DOI 10.1038/s41477-017-0035-3.
- Qian P., Song W., Zaizen-Iida M., Kume S., Wang G., Zhang Y., Kinoshita-Tsujimura K., Chai J., Kakimoto T. A Dof-CLE circuit controls phloem organization. *Nat. Plants*. 2022;8(7):817-827. DOI 10.1038/s41477-022-01176-0.
- Ren S.C., Song X.F., Chen W.Q., Lu R., Lucas W.J., Liu C.M. CLE25 peptide regulates phloem initiation in *Arabidopsis* through a CLERK-CLV2 receptor complex. *J. Integr. Plant Biol.* 2019;61(10):1043-1061. DOI 10.1111/jipb.12846.
- Robert H.S., Crhak Khaitova L., Mroue S., Benková E. The importance of localized auxin production for morphogenesis of reproductive organs and embryos in *Arabidopsis*. *J. Exp. Bot.* 2015;66(16):5029-5042. DOI 10.1093/jxb/erv256.

- Rodriguez-Villalon A. Wiring a plant: genetic networks for phloem formation in *Arabidopsis thaliana* roots. *New Phytol.* 2016;210(1): 45-50. DOI 10.1111/nph.13527.
- Rodriguez-Villalon A., Gujas B., van Wijk R., Munnik T., Hardtke C.S. Primary root protophloem differentiation requires balanced phosphatidylinositol-4,5-bisphosphate levels and systemically affects root branching. *Development.* 2015;142(8):1437-1446. DOI 10.1242/dev.118364.
- Ross-Elliott T.J., Jensen K.H., Haaning K.S., Wager B.M., Knoblauch J., Howell A.H., Mullendore D.L., Monteith A.G., Paultre D., Yan D., Otero S., Bourdon M., Sager R., Lee J.Y., Helariutta Y., Knoblauch M., Oparka K.J. Phloem unloading in *Arabidopsis* roots is convective and regulated by the phloem-pole pericycle. *eLife.* 2017;6:e24125. DOI 10.7554/eLife.24125.
- Roszak P., Heo J.O., Blob B., Toyokura K., Sugiyama Y., de Luis Balaguer M.A., Lau W.W.Y., Hamey F., Cirrone J., Madej E., Bouatta A.M., Wang X., Guichard M., Ursache R., Tavares H., Verstaen K., Wendrich J., Melnyk C.W., Oda Y., Shasha D., Ahnert S.E., Saeys Y., De Rybel B., Heidstra R., Scheres B., Grossmann G., Mähönen A.P., Denninger P., Göttgens B., Sozzani R., Birnbaum K.D., Helariutta Y. Cell-by-cell dissection of phloem development links a maturation gradient to cell specialization. *Science.* 2021;374(6575):eaba5531. DOI 10.1126/science.aba5531.
- Ruiz Sola M.A., Coiro M., Crivelli S., Zeeman S.C., Hansen S.S.K., Truernit E. *OCTOPUS-LIKE 2*, a novel player in *Arabidopsis* root and vascular development, reveals a key role for *OCTOPUS* family genes in root metaphloem sieve tube differentiation. *New Phytol.* 2017;216(4):1191-1204. DOI 10.1111/nph.14751.
- Růžička K., Ursache R., Hejátko J., Helariutta Y. Xylem development – from the cradle to the grave. *New Phytol.* 2015;207(3):519-535. DOI 10.1111/nph.13383.
- Sauer M., Balla J., Luschnig C., Wisniewska J., Reinöhl V., Friml J., Benková E. Canalization of auxin flow by Aux/IAA-ARF-dependent feedback regulation of PIN polarity. *Genes Dev.* 2006;20(20):2902-2911. DOI 10.1101/gad.390806.
- Scarpella E., Meijer A.H. Pattern formation in the vascular system of monocot and dicot plant species. *New Phytol.* 2004;164(2):209-242. DOI 10.1111/j.1469-8137.2004.01191.x.
- Scheres B., Wolkenfelt H., Willemsen V., Terlouw M., Lawson E., Dean C., Weisbeek P. Embryonic origin of the *Arabidopsis* primary root and root meristem initials. *Development.* 1994;120(9):2475-2487. DOI 10.1242/dev.120.9.2475.
- Schlereth A., Möller B., Liu W., Kientz M., Flipse J., Rademacher E.H., Schmid M., Jürgens G., Weijers D. MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. *Nature.* 2010;464(7290):913-916. DOI 10.1038/nature08836.
- Seo M., Kim H., Lee J.Y. Information on the move: vascular tissue development in space and time during postembryonic root growth. *Curr. Opin. Plant Biol.* 2020;57:110-117. DOI 10.1016/j.pbi.2020.08.002.
- Sjolund R.D. The phloem sieve element: a river runs through it. *Plant Cell.* 1997;9(7):1137-1146. DOI 10.1105/tpc.9.7.1137.
- Smet W., Sevilem I., de Luis Balaguer M.A., Wybouw B., Mor E., Miyashima S., Blob B., Roszak P., Jacobs T.B., Boekschoten M., Hooiveld G., Sozzani R., Helariutta Y., De Rybel B. DOF2.1 controls cytokinin-dependent vascular cell proliferation downstream of TMO5/LHW. *Curr. Biol.* 2019;29(3):520-529.e6. DOI 10.1016/j.cub.2018.12.041.
- Smetana O., Mäkilä R., Lyu M., Amiryousefi A., Sánchez Rodríguez F., Wu M.-F., Solé-Gil A., Leal Gavarrón M., Siligato R., Miyashima S., Roszak P., Blomster T., Reed J.W., Broholm S., Mähönen A.P. High levels of auxin signalling define the stem-cell organizer of the vascular cambium. *Nature.* 2019;565(7740):485-489. DOI 10.1038/s41586-018-0837-0.
- Smit M.E., Llavata-Peris C.I., Roosjen M., van Beijnum H., Novikova D., Levitsky V., Sevilem I., Roszak P., Slane D., Jürgens G., Mironova V., Brady S.M., Weijers D. Specification and regulation of vascular tissue identity in the *Arabidopsis* embryo. *Development.* 2020;147(8):dev186130. DOI 10.1242/dev.186130.
- Smit M.E., Weijers D. The role of auxin signaling in early embryo pattern formation. *Curr. Opin. Plant Biol.* 2015;28:99-105. DOI 10.1016/j.pbi.2015.10.001.
- Smith R.A., Schuetz M., Roach M., Mansfield S.D., Ellis B., Samuels L. Neighboring parenchyma cells contribute to *Arabidopsis* xylem lignification, while lignification of interfascicular fibers is cell autonomous. *Plant Cell.* 2013;25(10):3988-3999. DOI 10.1105/tpc.113.117176.
- Stadler R., Wright K.M., Lauterbach C., Amon G., Gahrz M., Feuerstein A., Oparka K.J., Sauer N. Expression of GFP-fusions in *Arabidopsis* companion cells reveals non-specific protein trafficking into sieve elements and identifies a novel post-phloem domain in roots. *Plant J.* 2005;41(2):319-331. DOI 10.1111/j.1365-313X.2004.02298.x.
- Tanaka H., Dhonukshe P., Brewer P.B., Friml J. Spatiotemporal asymmetric auxin distribution: a means to coordinate plant development. *Cell. Mol. Life Sci.* 2006;63(23):2738-2754. DOI 10.1007/s00018-006-6116-5.
- Truernit E. Sieve elements and their cell neighbours in the *Arabidopsis* root – roles and relationships. *J. Plant Physiol.* 2022;268:153569. DOI 10.1016/j.jplph.2021.153569.
- Vera-Sirera F., De Rybel B., Úrbez C., Kouklas E., Pesquera M., Álvarez-Mahecha J.C., Minguet E.G., Tuominen H., Carbonell J., Borst J.W., Weijers D., Blázquez M.A. A bHLH-based feedback loop restricts vascular cell proliferation in plants. *Dev. Cell.* 2015;35(4):432-443. DOI 10.1016/j.devcel.2015.10.022.
- Verma S., Attuluri V.P.S., Robert H.S. An essential function for auxin in embryo development. *Cold Spring Harb. Perspect. Biol.* 2021;13(4):a039966. DOI 10.1101/cshperspect.a039966.
- Wabnik K., Robert H.S., Smith R.S., Friml J. Modeling framework for the establishment of the apical-basal embryonic axis in plants. *Curr. Biol.* 2013;23(24):2513-2518. DOI 10.1016/j.cub.2013.10.038.
- Wallner E.S., López-Salmerón V., Belevich I., Poschet G., Jung I., Grünwald K., Sevilem I., Jokitalo E., Hell R., Helariutta Y., Agustí J., Lebovka I., Greb T. Strigolactone- and karrikin-independent SMXL proteins are central regulators of phloem formation. *Curr. Biol.* 2017;27(8):1241-1247. DOI 10.1016/j.cub.2017.03.014.
- Weijers D., Jürgens G. Auxin and embryo axis formation: the ends in sight? *Curr. Opin. Plant Biol.* 2005;8(1):32-37. DOI 10.1016/j.pbi.2004.11.001.
- Yang B., Minne M., Brunoni F., Plačková L., Petřík I., Sun Y., Nolf J., Smet W., Verstaen K., Wendrich J.R., Eekhout T., Hoyerová K., van Isterdael G., Haustraete J., Bishopp A., Farcot E., Novák O., Saeys Y., de Rybel B. Non-cell autonomous and spatiotemporal signalling from a tissue organizer orchestrates root vascular development. *Nat. Plants.* 2021;7(11):1485-1494. DOI 10.1038/s41477-021-01017-6.

**Благодарности.** Работа выполнена при поддержке бюджетного проекта FWN-2022-0020. Микроскопический анализ осуществлен в рамках проекта, поддержанного Российским научным фондом, грант № 20-14-00140. Авторы благодарны Центру коллективного пользования микроскопического анализа биологических объектов СО РАН (<http://www.bionet.nsc.ru/microscopu/>) за предоставленное оборудование. Авторы выражают глубокую признательность профессору Дольфу Вейерсу за любезно предоставленные репортерные линии *Arabidopsis thaliana*.

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 16.09.2022. После доработки 08.11.2022. Принята к публикации 10.11.2022.

Перевод на английский язык <https://vavilov.elpub.ru/jour>

## Компьютерный анализ особенностей регуляции гиперметилированных маркерных генов гепатокарциномы вирусными белками гепатита С

Е.А. Антропова<sup>1</sup>✉, Т.М. Хлебодарова<sup>1, 2</sup>, П.С. Деменков<sup>1, 2</sup>, А.С. Вензель<sup>1, 2</sup>, Н.В. Иванисенко<sup>1, 2</sup>, А.Д. Гавриленко<sup>1, 3</sup>, Т.В. Иванисенко<sup>1, 2</sup>, А.В. Адамовская<sup>2, 3</sup>, П.М. Ревва<sup>2, 3</sup>, И.Н. Лаврик<sup>4</sup>, В.А. Иванисенко<sup>1, 2, 3</sup>

<sup>1</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>2</sup> Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия

<sup>3</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

<sup>4</sup> Магдебургский университет им. Отто фон Герике, медицинский факультет, Магдебург, Германия

✉ nzhenia@bionet.nsc.ru

**Аннотация.** Вирус гепатита С (ВГС) считается фактором риска для возникновения гепатоцеллюлярной карциномы (ГЦК). Известно, что большую роль в молекулярно-генетических механизмах вирус-индуцированного онкогенеза играют эпигенетические изменения. Аберрантное метилирование ДНК служит медиатором эпигенетических изменений, которые тесно связаны с патогенезом ГЦК, и признано биомаркером для его ранней диагностики. С помощью ANDSystem проведены реконструкция и оценка статистической значимости путей потенциальной регуляции вирусными белками ВГС 32 генов человека, гиперметилированных при ГЦК. Среди исследованных генов были как онкосупрессоры, так и проопухолевые гены, идентифицированных по данным полногеномного анализа метилирования ДНК. Реконструированы регуляторные пути, включающие белок-белковые взаимодействия, регуляцию экспрессии генов, регуляцию активности, стабильности и транспорта белков. Среди статистически значимых оказались пути регуляции экспрессии. Показано, что восемь из десяти белков ВГС являются участниками данных путей. Белок ВГС NS3 был вовлечен в наибольшее число регуляторных путей. NS3 связан с регуляцией пяти генов-онкосупрессоров, что может свидетельствовать о его центральной роли в патогенезе ГЦК. Анализ реконструированных путей показал, что при ингибировании транскрипционных факторов в результате связывания с вирусными белками, экспрессия ряда онкосупрессоров (*WT1*, *MGMT*, *SOCST1*, *P53*) подавлялась, тогда как экспрессия других (*RASF1*, *RUNX3*, *WIF1*, *DAPK1*) активировалась. Таким образом, с помощью реконструкции генных сетей показано, что вирусные белки гепатита С способны влиять не только на статус метилирования генов-онкосупрессоров, но и на их транскрипционную регуляцию. Полученные результаты могут быть использованы при поиске фармакологических мишеней для разработки новых средств против ГЦК, индуцированной ВГС.

Ключевые слова: гепатоцеллюлярная карцинома; вирус гепатита С; регуляция экспрессии; гиперметилирование; регуляторные пути; генные сети; биоинформатика.

**Для цитирования:** Антропова Е.А., Хлебодарова Т.М., Деменков П.С., Вензель А.С., Иванисенко Н.В., Гавриленко А.Д., Иванисенко Т.В., Адамовская А.В., Ревва П.М., Лаврик И.Н., Иванисенко В.А. Компьютерный анализ особенностей регуляции гиперметилированных маркерных генов гепатокарциномы вирусными белками гепатита С. *Вавиловский журнал генетики и селекции*. 2022;26(8):733-742. DOI 10.18699/VJGB-22-89

## Computer analysis of regulation of hepatocarcinoma marker genes hypermethylated by HCV proteins

Е.А. Antropova<sup>1</sup>✉, T.M. Khlebodarova<sup>1, 2</sup>, P.S. Demenkov<sup>1, 2</sup>, A.S. Venzel<sup>1, 2</sup>, N.V. Ivanisenko<sup>1, 2</sup>, A.D. Gavrilenko<sup>1, 3</sup>, T.V. Ivanisenko<sup>1, 2</sup>, A.V. Adamovskaya<sup>2, 3</sup>, P.M. Revva<sup>2, 3</sup>, I.N. Lavrik<sup>4</sup>, V.A. Ivanisenko<sup>1, 2, 3</sup>

<sup>1</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>2</sup> Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>4</sup> Translational Inflammation Research, Medical Faculty, Otto von Guericke University Magdeburg, Magdeburg, Germany

✉ nzhenia@bionet.nsc.ru

**Abstract.** Hepatitis C virus (HCV) is a risk factor that leads to hepatocellular carcinoma (HCC) development. Epigenetic changes are known to play an important role in the molecular genetic mechanisms of virus-induced oncogenesis. Aberrant DNA methylation is a mediator of epigenetic changes that are closely associated with the HCC pathogenesis and considered a biomarker for its early diagnosis. The ANDSystem software package was used to reconstruct and evaluate the statistical significance of the pathways HCV could potentially use to regulate 32 hypermethylated genes in HCC, including both oncosuppressor and protumorigenic ones identified by genome-wide analysis of DNA methyla-

tion. The reconstructed pathways included those affecting protein-protein interactions (PPI), gene expression, protein activity, stability, and transport regulations, the expression regulation pathways being statistically significant. It has been shown that 8 out of 10 HCV proteins were involved in these pathways, the HCV NS3 protein being implicated in the largest number of regulatory pathways. NS3 was associated with the regulation of 5 tumor-suppressor genes, which may be the evidence of its central role in HCC pathogenesis. Analysis of the reconstructed pathways has demonstrated that following the transcription factor inhibition caused by binding to viral proteins, the expression of a number of oncosuppressors (*WT1*, *MGMT*, *SOCS1*, *P53*) was suppressed, while the expression of others (*RAS1*, *RUNX3*, *WIF1*, *DAPK1*) was activated. Thus, the performed gene-network reconstruction has shown that HCV proteins can influence not only the methylation status of oncosuppressor genes, but also their transcriptional regulation. The results obtained can be used in the search for pharmacological targets to develop new drugs against HCV-induced HCC.

Key words: hepatocellular carcinoma; hepatitis C virus; expression regulation; methylation; regulatory pathways; gene networks; bioinformatics.

**For citation:** Antropova E.A., Khlebodarova T.M., Demenkov P.S., Venzel A.S., Ivanisenko N.V., Gavrilenko A.D., Ivanisenko T.V., Adamovskaya A.V., Revva P.M., Lavrik I.N., Ivanisenko V.A. Computer analysis of regulation of hepatocarcinoma marker genes hypermethylated by HCV proteins. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):733-742. DOI 10.18699/VJGB-22-89

## Введение

По статистике за 2020 г. рак печени признан третьей по значимости причиной смерти от рака в мире. За этот же год в мире было зарегистрировано более 900 тыс. новых случаев этой патологии (International Agency for Research on Cancer, <https://gco.iarc.fr/today/home>). Доминирующим типом первичного рака печени (~90 % случаев) является гепатоцеллюлярная карцинома (ГЦК) (Llovet et al., 2016). Развитие ГЦК может быть вызвано действием факторов риска, таких как воздействие афлатоксина, употребление алкоголя, заражение вирусами гепатита В или С (ВГС), цирроз печени, неалкогольная жировая болезнь печени, неалкогольный стеатогепатит, метаболический синдром, ожирение, диабет II типа и генетическая предрасположенность (McGlynn et al., 2021).

В настоящее время накоплено множество данных по ассоциации ВГС с нарушениями функций печени, развитием цирроза и ГЦК (Rabaan et al., 2020). Проникая в организм, ВГС стремится осуществлять контроль над биологическими процессами, протекающими в клетках хозяина, с целью повышения своей выживаемости и эффективности репликации. Более чем у 70 % первично инфицированных заболевание принимает хроническое течение, при котором у больных наблюдается прогрессирующий фиброз и цирроз печени на фоне ее длительного воспаления (Jaroszewicz et al., 2015). Используя различные механизмы для кооптации инфицированных клеток, вирус может непреднамеренно привести к развитию ГЦК (D'souza et al., 2020). Однако молекулярно-генетические механизмы вирус-индуцированного канцерогенеза до сих пор плохо изучены.

Кроме того, патогенез ГЦК связан с эпигенетическими модификациями, а aberrантное метилирование ДНК является медиатором эпигенетических изменений (Fernández-Barrena et al., 2020) и может служить биомаркером для ранней диагностики ГЦК (Zhang C. et al., 2016; Xu et al., 2017).

Для установления функциональных связей между генами и выяснения молекулярных механизмов биологических процессов широкое применение нашли методы реконструкции генных сетей. Ранее нами была разработана программно-информационная система ANDSystem, пред-

назначенная для реконструкции генных сетей на основе знаний, экстрагированных из фактографических баз данных и извлеченных с помощью методов текст-майнинга из научных публикаций (Ivanisenko V.A. et al., 2015, 2019; Ivanisenko T.V. et al., 2020). С применением ANDSystem были реконструированы молекулярные механизмы ряда патологий: преэклампсия (Glotov et al., 2015), туберкулез (Bragina et al., 2016), коморбидные состояния астмы и гипертонии (Saik et al., 2018), COVID-19 (Ivanisenko N.V. et al., 2020), жизненный цикл ВГС (Saik et al., 2016) и др.

В настоящей работе с использованием ANDSystem проведена реконструкция регуляторных путей, описывающих потенциальные механизмы модуляции гиперметилированных при ГЦК генов белками ВГС. В анализе были рассмотрены 32 гена, известных как гиперметилированные маркеры ГЦК. Среди семи типов реконструированных регуляторных путей, включающих белок-белковые взаимодействия, регуляцию экспрессии генов, регуляцию активности, стабильности и транспорта белков, статистически значимыми оказались пути регуляции экспрессии генов. Выявлено девять генов-маркеров, которые потенциально могут быть подвержены регуляции со стороны белков ВГС. Среди них три гена-супрессора ГЦК (*MGMT*, *SOCS1* и *TP53*) могут быть подвержены негативной регуляции, а ген-супрессор апоптоза (*TERT*) – положительной регуляции.

## Материалы и методы

**Гены, гиперметилированные при ГЦК.** Информация о генах, находящихся в гиперметилированном состоянии при ГЦК, была взята из опубликованных статей (табл. 1). Рассматривались только те гены, гиперметилирование которых ассоциировано с ГЦК, что было показано с помощью анализа и метаанализа в опубликованных работах. Схематичное описание алгоритма обработки данных представлено на рис. 1.

**Реконструкция регуляторных путей** проводилась с помощью программно-информационной системы ANDSystem (Ivanisenko V.A. et al., 2019). Данная система была разработана для реконструкции генных сетей на основе автоматического анализа текстов научных публикаций и фактографических баз данных. ANDSystem

**Таблица 1.** Список гиперметилированных генов, использованных в анализе

Ген	Кодируемый белок	Литературный источник
<i>APC</i>	Adenomatous polyposis coli	Zhang C. et al., 2016
<i>COL7A1</i>	Collagen alpha-1(VII) chain	Schulze et al., 2015
<i>COL16A1</i>	Collagen alpha-1(XVI) chain	»
<i>DAPK1</i>	Death-associated protein kinase 1	Zhang C. et al., 2016
<i>DSE</i>	Dermatan-sulfate epimerase	Cheng et al., 2018
<i>FAM55C</i>	NXPE family member 3	»
<i>FAT4</i>	Protocadherin Fat 4	Schulze et al., 2015
<i>GALNT3</i>	Polypeptide N-acetylgalactosaminyltransferase 3	Cheng et al., 2018
<i>GSTP1</i>	Glutathione S-transferase P	Zhang C. et al., 2016
<i>IGFALS</i>	Insulin-like growth-factor-binding protein, acid labile subunit	Neumann et al., 2012
<i>KCNA3</i>	Potassium voltage-gated channel subfamily A member 3	Hernandez-Meza et al., 2021
<i>LDHB</i>	L-lactate dehydrogenase B chain	»
<i>MGMT</i>	O6-methylguanine-DNA methyltransferase	Zhang C. et al., 2016
<i>NEBL</i>	Nebulette	Cheng et al., 2018
<i>NEFH</i>	Neurofilament heavy polypeptide	Revill et al., 2013
<i>OPCML</i>	Opioid binding protein/cell adhesion molecule-like	Zhang C. et al., 2016
<i>TP53</i>	Cellular tumor antigen p53	»
<i>PER3</i>	Period circadian protein homolog 3	Neumann et al., 2012
<i>PRDM2</i>	PR domain containing 2	Zhang C. et al., 2016
<i>PROZ</i>	Vitamin K-dependent protein Z	Neumann et al., 2012
<i>RARβ</i>	Retinoic acid receptor beta	Zhang C. et al., 2016
<i>RASSF1A</i>	Ras association domain-containing protein 1	»
<i>RUNX3</i>	Runt-related transcription factor 3	»
<i>SFRP1</i>	Secreted frizzled-related protein 1	»
<i>SMPD3</i>	Sphingomyelin phosphodiesterase 3	Revill et al., 2013
<i>SOCS1</i>	Suppressor of the cytokine signalling 1	Zhang C. et al., 2016
<i>SPINT2</i>	Serine peptidase inhibitor, Kunitz type, 2	Zhang C. et al., 2016; Hernandez-Meza et al., 2021
<i>TERT</i>	Telomerase reverse transcriptase	Zhang H. et al., 2015
<i>TSPYL5</i>	Testis-specific Y-encoded-like protein 5	Hernandez-Meza et al., 2021
<i>TTC36</i>	Tetratricopeptide repeat protein 36	Jing et al., 2022
<i>WIF1</i>	WNT inhibitory factor 1	Zhang C. et al., 2016
<i>WT1</i>	Wilms tumor 1	»

включает в себя базу знаний, содержащую более 40 млн фактов о молекулярно-генетических взаимодействиях, включая физические межмолекулярные взаимодействия, регуляцию экспрессии генов, регуляцию активности, стабильности и транспорта белков. Работа над реконструкцией и анализом генных сетей в ANDSystem осуществляется в программе ANDVisio. Для реконструкции регуляторных путей использовалась функция Pathway Wizard, реализованная в ANDVisio, которая по заданному шаблону выполняет поисковые обращения к базе знаний. Схематическое описание шаблонов приведено в табл. 2.

Например, шаблон P4 означает поиск всех возможных молекулярно-генетических путей в базе знаний ANDSystem, удовлетворяющих следующему требованию: первым

участником пути является вирусный белок (Vp); вторым участником – белок человека (Hr); третьим – ген человека из списка генов-мишеней (Tg); последним – белок (Tr), кодируемый Tg. Далее по тексту в качестве генов-мишеней будут рассматриваться гены-маркеры ГЦК. Взаимодействия между участниками пути представлены следующими типами: Vp и Hr связаны белок-белковыми взаимодействиями (PPI); Hr и Tg – взаимодействием по типу «регуляция экспрессии» (*Exp reg*), где Hr – регулятор экспрессии гена Tg; Tg и Tr – взаимодействием по типу «экспрессия» (*Exp*), т. е. белок Tr является продуктом экспрессии гена Tg. Примеры реконструкции регуляторных путей в ANDSystem с использованием шаблонов представлены в предыдущей работе (Ivanisenko V.A. et al., 2022).

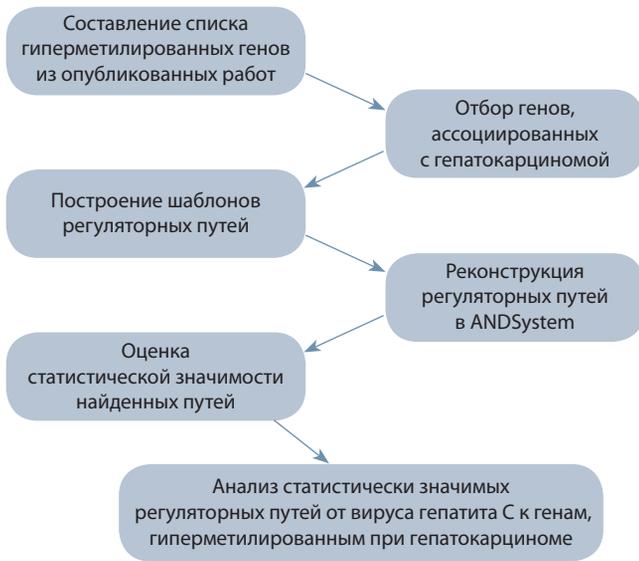


Рис. 1. Схематичное описание алгоритма обработки данных.

**Оценка статистической значимости регуляторных путей.** С помощью шаблонов из табл. 2 рассчитывалось число генов-маркеров  $K$ , являющихся участниками регуляторных путей, а также число таких участников из выборки контрольных генов. Вероятность наблюдения по случайным причинам числа  $K$  оценивалась с помощью стандартного гипергеометрического распределения с использованием функции `hypergeom` из пакета SciPy 1.8.0 (<https://scipy.org>). Для статистической обработки в качестве контроля была взята группа генов, предложенная в работе (Hoshida et al., 2008) для предсказания исхода заболевания ГЦК по уровню их экспрессии.

## Результаты

### Реконструкция потенциальных регуляторных путей влияния белков ВГС на гены-маркеры ГЦК

На начальном этапе исследования была собрана группа гиперметилированных генов-маркеров ГЦК (см. табл. 1). Список был сформирован на основе опубликованных данных полногеномного анализа метилирования ДНК и включал 30 генов, экспрессия которых по данным проведенных исследований была снижена при гепатоцеллюлярной карциноме, и два гена (*WT1* и *TERT*) с повышенной экспрессией.

Для реконструкции регуляторных путей использовалась программно-информационная система ANDSystem. Поиск запросы к базе знаний ANDSystem основывались на шаблонах регуляторных путей, представленных в табл. 2. Шаблоны описывают различные типы регуляторных путей, определяемые разными комбинациями молекулярно-генетических взаимодействий, включая белок-белковые взаимодействия, регуляцию экспрессии генов, регуляцию активности, стабильности и транспорта белков.

Анализ статистической значимости регуляторных путей, автоматически реконструированных ANDSystem по заданным шаблонам, показал, что среди семи проанали-

Таблица 2. Шаблоны для поиска регуляторных путей, описывающих модуляцию вирусными белками генов-маркеров ГЦК

N шаблона	Схема шаблона
P1	$Vp \xrightarrow{PPI} Tr$
P2	$Vp \xrightarrow{PPI} Hr \xrightarrow{PPI} Tr$
P3	$Vp \xrightarrow{PPI} Hr \xrightarrow{Act/Stab/Pr/PPM/Tr} Tr$
P4	$Vp \xrightarrow{PPI} Hr \xrightarrow{Exp\ reg} Tg \xrightarrow{Exp} Tr$
P5	$Vp \xrightarrow{PPI} Hr \xrightarrow{Exp\ reg} Hg \xrightarrow{Exp} Hr \xrightarrow{Exp\ reg} Tg \xrightarrow{Exp} Tr$
P6	$Vp \xrightarrow{PPI} Hr \xrightarrow{Exp\ reg} Hg \xrightarrow{Exp} Hr \xrightarrow{Act/Stab/Pr/PPM/Tr} Tr$
P7	$Vp \xrightarrow{PPI} Hr \xrightarrow{Exp\ reg} Hg \xrightarrow{Exp} Hr \xrightarrow{PPI} Tr$

Примечание. Vp – белки вируса гепатита С; Hr – любые белки человека, участвующие во взаимодействиях; Hg – любые гены человека, участвующие во взаимодействиях; Tg – гены-мишени (гены-маркеры гепатокарциномы); Tr – белки-мишени (кодируемые Tg); PPI – белок-белковые взаимодействия; Act/Stab/Pr/PPM/Tr – регуляция активности или стабильности, или протеолиз, или посттрансляционные модификации, или транспорт; Exp reg – регуляция экспрессии генов; Exp – экспрессия генов (наработка белков).

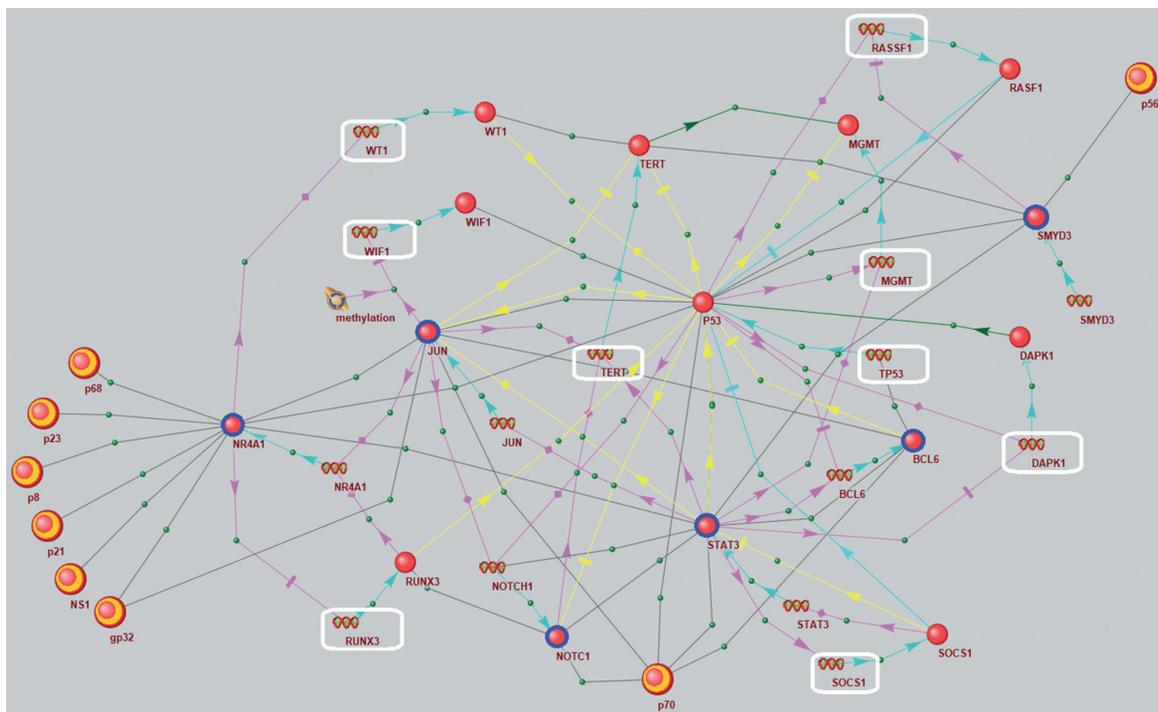
Таблица 3. Результаты оценки значимости регуляторных путей, описываемых разными шаблонами

Шаблон регуляторного пути	Количество генов-маркеров, участников регуляторного пути	P-val	FDR
P1	0	–	–
P2	15	0.52	0.62
P3	5	0.17	0.34
P4	9	<b>0.0054</b>	<b>0.032</b>
P5	21	0.047	0.14
P6	10	0.39	0.58
P7	23	0.83	0.83

Примечание. P-val – уровень статистической значимости; FDR – уровень статистической значимости с учетом множественного сравнения по методу False discovery rate (ожидаемая доля ложных отклонений).

зированных типов значимыми оказались пути регуляции экспрессии (шаблон P4 в табл. 3). Данный шаблон описывает регуляторные пути, включающие четыре участника: 1) вирусные белки; 2) транскрипционные факторы человека, вовлеченные в белок-белковые взаимодействия с вирусными белками; 3) гены-маркеры, представленные в табл. 1, экспрессия которых регулируется транскрипционными факторами-участниками под номером 2; 4) белки-продукты генов-маркеров.

Генная сеть, описывающая пути регуляции генов-маркеров ГЦК белками ВГС, включает восемь белков вируса гепатита С, семь промежуточных белков-посредников из числа белков хозяина, участвующих в белок-белковых взаимодействиях с белками ВГС, и девять генов (*DAPK1*, *SOC31*, *MGMT*, *RASSF1*, *RUNX3*, *TP53*, *WIF1*, *WT1*, *TERT*), aberrantная экспрессия которых коррелирует с прогрессией ГЦК (рис. 2).



**Рис. 2.** Генная сеть, включающая статистически значимые регуляторные пути влияния вирусных белков на экспрессию генов-маркеров ГЦК, реконструированная с помощью ANDSysystem по шаблону P4.

Белки вируса гепатита С (желто-красные большие шарики) – p8 (Non-structural protein 4A, NS4A), p21 (Core, Capsid protein C), p23 (Protease NS2-3), gp32 (Envelope glycoprotein E1), NS1 (Envelope glycoprotein E2), p56 (NS5A), p68 (NS5B), p70 (Hepacivirin, NS3); промежуточные посредники (сине-красные шарики): BCL6 (B-cell lymphoma 6 protein), NOTC1 (neurogenic locus notch homolog protein 1), NR4A1 (Nuclear receptor subfamily 4 group A member 1), JUN (c-Jun/activator protein 1), SMYD3 (lysine methyltransferase SET and MYND domain containing protein 3), STAT3 (Signal transducer and activator of transcription 3); гиперметилированные гены (выделены белыми рамками) и их белковые продукты: DAPK1 (Death associated protein kinase 1), MGMT (Methylated-DNA-protein-cysteine methyltransferase), RASSF1 (Ras association domain family member 1), RUNX3 (Runt-related transcription factor 3), SOCS1 (Suppressor of cytokine signaling 1), TERT (telomerase reverse transcriptase), TP53 (Tumor protein p53), WIF1 (Wnt inhibitory factor 1), WT1 (Wilms tumor protein).

### Анализ регуляторных путей влияния белков ВГС на гены-маркеры ГЦК

В регуляторных путях оказались задействованы восемь из десяти белков ВГС и шесть генов человека, являющихся промежуточными участниками, с которыми вирусные белки могут образовывать белковые гетерокомплексы. Среди последних транскрипционные факторы: STAT3 (Signal transducer and activator of transcription 3), NR4A1 (Nuclear receptor subfamily 4 group A member 1), JUN (c-Jun/activator protein 1), BCL6 (B-cell lymphoma 6 protein), а также трансмембранный рецептор NOTC1 (neurogenic locus notch homolog protein 1) и гистонметилтрансферазы SMYD3 (lysine methyltransferase SET and MYND domain containing protein 3).

Наибольшее число вирусных белков было связано с регуляцией *RUNX3* и *WT1*. Шесть вирусных белков (NS4A, Core, p23, gp32, NS1 и NS5B) взаимодействовали с NR4A1, который является общим регулятором экспрессии этих двух генов-маркеров ГЦК.

Если анализировать взаимодействия вирусных белков с клеточными регуляторами экспрессии, то можно увидеть, что белок NS3 (p70) потенциально может регулировать экспрессию пяти генов-супрессоров опухоли и TERT.

Рассмотрим возможности реализации этих регуляторных путей более подробно.

### p8, p21, p68, gp32, p23, NS1/NR4A1/RUNX3, WT1.

Данный регуляторный путь предполагает возможность влияния шести вирусных белков (p8, p21, p68, gp32, p23, NS1) ВГС на развитие гепатокарциномы через контроль активности генов *RUNX3* и *WT1* транскрипционным фактором NR4A1. Действительно, транскрипционный фактор NR4A1 напрямую взаимодействует с промоторами генов *RUNX3* и *WT1*, супрессируя активность *RUNX3* и активируя *WT1* (Nowyhed et al., 2015; Zong et al., 2017). Оба фактора участвуют в регуляции апоптоза. Так, *RUNX3* способствует активации внешнего, TRAIL-индуцированного пути апоптоза (Kim et al., 2019), а *WT1* контролирует митохондриальный (внутренний) путь апоптоза через регуляцию гена антиапоптотического белка *Bcl-2*, причем в зависимости от типа клеток оказывает на экспрессию гена *Bcl-2* как позитивное, так и негативное действие (Maуо et al., 1999; Loeb, 2006). Показано, что при ГЦК наблюдается повышенная экспрессия гена *WT1*, которая обусловлена гиперметилированием его промотора и коррелирует с плохим прогнозом заболевания (Sera et al., 2008; Mžik et al., 2016). Эти данные позволяют предположить, что роль *WT1* в прогрессии ГЦК связана с блокированием апоптоза.

Экспериментально показано, что коровый белок вируса гепатита С ингибирует экспрессию генов *NR4A1* и

*RUNX3* в инфицированных клетках (Tan, Li, 2015), способствуя подавлению внешнего пути апоптоза. С помощью Y2H теста (two hybrid test) выявлена возможность взаимодействия белка NR4A1 с вирусными белками CORE, E1, E2, NS2, NS4A и NS5B (de Chassey et al., 2008), однако, за исключением CORE, эффекты других белков ВГС на активность ТФ не исследованы.

**E1, NS3, Core, p23, NS1, p68/JUN, NOTC1, STAT3/TERT.** Абберрантная экспрессия гена *TERT*, связанная в том числе с гиперметилированием его промотора, является прогностическим маркером ГЦК (Zhang H. et al., 2015; Zucman-Rossi et al., 2015; Oversoe et al., 2020). *TERT* влияет на прогрессию заболевания через стимуляцию клеточной пролиферации благодаря реактивации экспрессии его гена в клетках карциномы (Nault et al., 2019; In der Stroth et al., 2020). Показано, что в клетках, инфицированных ВГС, активность *TERT* тоже повышена, частично через прямое взаимодействие корового белка с ферментом (Zhu et al., 2010, 2017), однако в целом механизмы влияния белков вируса гепатита С на активность *TERT* не ясны. Данный регуляторный путь предполагает возможность действия белков NS3, Core, E1, p23, NS1 и NS5B вируса гепатита С на экспрессию гена *TERT* через взаимодействие с белками JUN (AP-1), STAT3, NOTC1.

Действительно, экспериментально показано, что существует возможность влияния на экспрессию *TERT* через транскрипционные факторы AP-1 и STAT3, поскольку они являются его прямыми регуляторами (Konnikova et al., 2005; Takakura et al., 2005), а также через NOTC1 сигнальный путь (Sawangarun et al., 2018). Более того, белок NS3 вируса гепатита С влияет на активность NOTC1 через транскрипционный фактор SRCAP (Iwai et al., 2011), а также на экспрессию AP1- и STAT3-регулируемых генов (Hassan et al., 2005, 2007; Machida et al., 2006; Li et al., 2010), однако конкретные механизмы реализации этих влияний в инфицированных клетках гепатоцитов практически не изучены.

**gp32, p70/JUN/WIF1.** Этот регуляторный путь описывает влияние белков NS3 и E1 вируса гепатита С на экспрессию гена *WIF1* (Wnt inhibitory factor 1) через взаимодействие с ТФ c-Jun/AP-1. *WIF1* является онкосупрессором, способствует снижению роста клеток при ГЦК (Deng et al., 2010), и уровень экспрессии его гена может быть прогностическим признаком течения заболевания (Huang et al., 2011).

Экспериментально показано, что существует как возможность прямого действия белков NS3 и E1 вируса гепатита С на активность c-Jun/AP-1 (de Chassey et al., 2008), так и действие последнего на экспрессию гена *WIF1* через взаимодействие с DNMT1 метилтрансферазой (DNA methyltransferase 1), которое ведет к супрессии *WIF1* в раковых клетках желчного пузыря (Lin et al., 2018). Однако какие механизмы супрессии гена *WIF1* реализуются в клетках гепатокарциномы, инфицированных ВГС, неизвестно.

**p70/STAT3/MGMT, DAPK1, SOCS1.** Этот регуляторный путь инициируется неструктурным белком NS3 (p70) ВГС, влияющим на активность генов *MGMT*, *SOCS* и *DAPK1* через взаимодействие с ТФ STAT3. Белки DAPK1 (Death-associated protein kinase 1), MGMT (Methylated-

DNA-protein-cysteine methyltransferase) и SOCS1 (Suppressor of cytokine signaling 1) считаются супрессорами опухолей, их низкая экспрессия в карциномах коррелирует с прогрессией заболевания (Gui et al., 2011; Jiang et al., 2019; Chen J. et al., 2020; Chen P. et al., 2020; Song et al., 2020).

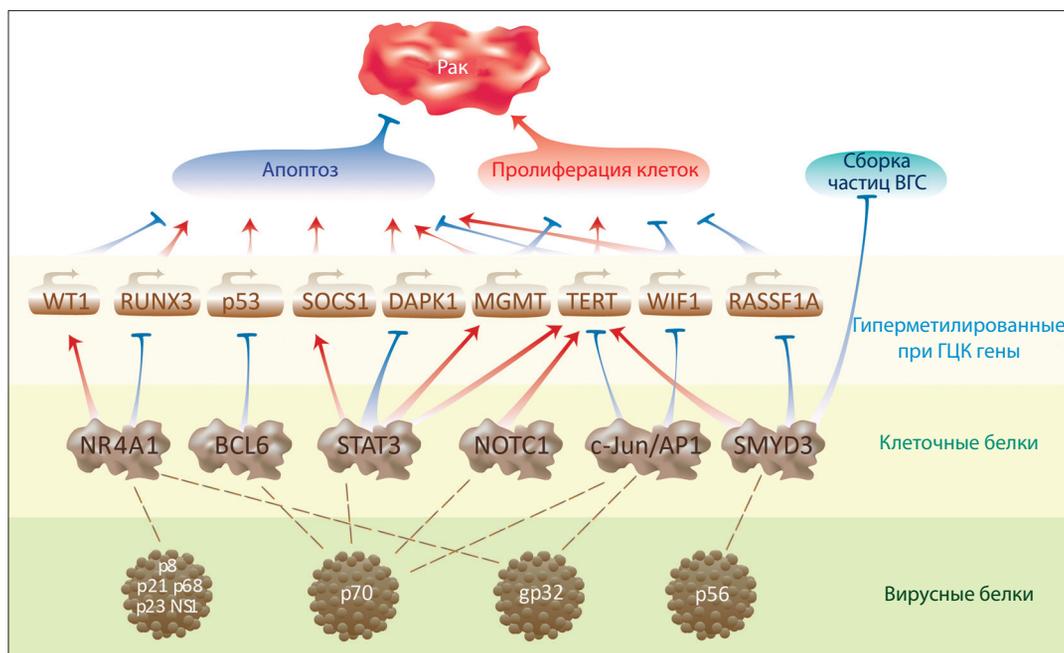
Экспериментально показано, что NS3 может напрямую взаимодействовать с ТФ STAT3 (de Chassey et al., 2008), который участвует в регуляции экспрессии *MGMT*, *SOCS* и *DAPK1* (Kohsaka et al., 2012; Benderska, Schneider-Stock, 2014; Yang C. et al., 2015), однако влияние транскрипционного фактора STAT3 на экспрессию данных генов неоднозначно и может быть связано с клеточной специализацией. Что касается механизмов регуляции экспрессии этих генов в клетках гепатокарциномы, инфицированных ВГС, то они не исследованы.

**p70/BCL6/TP53.** *TP53* – ключевой активатор внутреннего пути апоптоза. Вирусный белок NS3 (p70) оказывает влияние на *TP53* через взаимодействие с транскрипционным фактором BCL6 (B-cell lymphoma 6 protein). *TP53* признан маркером ГЦК, его низкая экспрессия коррелирует с плохим прогнозом заболевания (Liu et al., 2012; Ye et al., 2017). Транскрипционный фактор BCL6 является репрессором гена *TP53* в клетках лимфоидного ряда. Его конститутивная экспрессия защищает В-лимфоциты от апоптоза, вызванного повреждением ДНК (Phan, Dallafavera, 2004; Jardin et al., 2007). Данные о влиянии вируса гепатита С на экспрессию и индукцию мутаций в генах *TP53* и *BCL6* в этих клетках противоречивы (Machida et al., 2004; Tucci et al., 2013). Взаимодействие NS3 белка ВГС с транскрипционным фактором BCL6 обсуждается в работе (Han et al., 2016), но конкретные механизмы влияния NS3 на активность транскрипционного фактора не исследованы.

## Обсуждение

Исследуемый набор гиперметилированных при ГЦК генов-маркеров этого заболевания (см. табл. 1) включал 30 генов с пониженной экспрессией и два гена с повышенной экспрессией. С использованием ANDSystem были реконструированы регуляторные пути, посредством которых белки ВГС способны влиять на экспрессию этих генов-маркеров. Схема взаимосвязи вирусных и клеточных белков, гиперметилированных генов и ключевых биологических процессов, ассоциированных с ГЦК, показана на рис. 3. Согласно данным литературы, гены *WT1*, *RUNX3*, *TP53* и *SOCS1* тесно связаны с апоптозом (Mayo et al., 1999; Loeb, 2006; Kim et al., 2019), а гены *MGMT*, *TERT*, *RASSF1A* и *WIF1* – как с апоптозом, так и с клеточной пролиферацией (He et al., 2005; Sarin et al., 2005; Choi et al., 2008; Feng et al., 2014; Chen J. et al., 2020; Ni et al., 2020).

Проведенный анализ показал потенциальную возможность реализации выявленных регуляторных путей влияния белков вируса гепатита С на активность генов-маркеров ГЦК, однако эффекты, которые оказывают белки ВГС на функцию регуляторных белков человека при образовании белок-белковых комплексов с ними, в настоящее время плохо изучены. К сожалению, это не позволяет однозначно интерпретировать реконструированные регуляторные пути. От того, останется ли способность регули-



**Рис. 3.** Связь реконструированных регуляторных путей с ключевыми биологическими процессами, ассоциированными с ГЦК.

ровать экспрессию генов у регуляторного белка организма хозяина после взаимодействия с вирусным белком или нет, зависит, будет ли регуляторный путь функционировать как активатор или как супрессор экспрессии целевого гена. Изучение этих эффектов требует проведения дальнейших экспериментальных исследований и компьютерного молекулярного моделирования.

В литературе описываются эффекты, которые могут оказывать вирусные белки на функцию хозяйских белков в результате взаимодействия с ними. В частности, белок NS5A связывается с SMYD3 в цитоплазме и ингибирует транслокацию SMYD3 в ядро (Chen M. et al., 2016). Если предположить, что регуляторные белки организма хозяина теряют способность осуществлять регуляцию экспрессии генов вследствие образования комплексов с вирусными белками, то можно ожидать следующих эффектов. При рассмотрении путей регуляции экспрессии онкосупрессоров из семи путей четыре (обеспечивающие ингибирование *RASFI*, *RUNX3*, *WIF1* и *DAPK1*) будут подавлены, что может привести к их активации белками ВГС, что, в свою очередь, будет препятствовать канцерогенезу. В оставшихся трех путях будет подавлена активация *MGMT*, *SOCS1* и *TP53*, что, вероятно, может оказать проопухолевый эффект. В представленных путях (см. рис. 2) *WT1*, *MGMT*, *SOCS1* и *TP53* активируются соответствующими факторами (Phan et al., 2004; Kohsaka et al., 2012; Yang S. et al., 2015; Zong et al., 2017), а экспрессия генов *RASFI*, *RUNX3*, *WIF1* и *DAPK1* контролируется негативно (Guo et al., 2011; Benderska, Schneider-Stock, 2014; Nowyhed et al., 2015; Lin et al., 2018).

В случае *TERT* и *WT1*, которые относятся к генам с проопухолевой активностью, можно ожидать подавления гена *WT1*, что приведет к негативному эффекту для канцерогенеза. Что касается *TERT*, который участвует в подавлении

апоптоза и стимулирует клеточную пролиферацию (Nault et al., 2019; In der Stroth et al., 2020), то, согласно нашим результатам (см. рис. 2), он оказался под управлением трех разных регуляторных путей. Экспрессия *TERT* активировалась двумя путями с участием хозяйских STAT3 и NOTC1 (Konnikova et al., 2005; Sawangarun et al., 2018). Один путь вел к подавлению экспрессии с участием c-JUN/AP-1 (Takakura et al., 2005). Взаимодействие вирусных белков с белками STAT3, NOTC1 и c-JUN/AP-1 способно привести к блокировке данных регуляторных путей. Особый интерес представляет путь c-JUN/AP-1, поскольку ингибирование этого транскрипционного фактора вирусными белками (gp32 и NS3) потенциально способствует активации *TERT*. Сделанные предположения хорошо согласуются с данными по дифференциальной экспрессии генов при острой инфекции гепатитом С, согласно которым в инфицированных клетках наблюдалась повышенная экспрессия гена *TERT* (Paric et al., 2012). Описанный путь представляет собой перспективную фармакологическую мишень.

Таким образом, в рамках рассмотренных предположений наблюдается разнонаправленная регуляция экспрессии генов маркеров ГЦК. Это может свидетельствовать в пользу того, что не все регуляторные пути, контролируемые вирусными белками, могут быть отнесены к факторам риска ГЦК. Однако регуляторные пути, обеспечивающие проопухолевую активность вирусных белков, несомненно, заслуживают дополнительного изучения для понимания механизмов вирус-индуцированного канцерогенеза ГЦК. В частности, подавление вирусными белками экспрессии ген-супрессоров опухолей способно усиливать эффект их метилирования при ГЦК или имитировать этот эффект в случае, когда эти гены не метилированы, и спровоцировать возникновение ГЦК или осложнить его течение.

## Заклучение

С использованием компьютерных методов реконструкции генных сетей, представленных в программе ANDSystem, установлены статистически значимые пути генетической регуляции маркеров ГЦК белками ВГС. Полученные результаты описывают потенциальные механизмы участия белков ВГС в патогенезе ГЦК и могут быть полезны при планировании экспериментальных исследований для поиска новых мишеней для создания лекарств и профилактических средств, снижающих риск развития ГЦК при заражении ВГС.

## Список литературы / References

- Benderska N., Schneider-Stock R. Transcription control of DAPK. *Apoptosis*. 2014;19(2):298-305. DOI 10.1007/s10495-013-0931-6.
- Bragina E.Y., Tiys E.S., Rudko A.A., Ivanisenko V.A., Freidin M.B. Novel tuberculosis susceptibility candidate genes revealed by the reconstruction and analysis of associative networks. *Infect. Genet. Evol.* 2016;46:118-123. DOI 10.1016/j.meegid.2016.10.030.
- Chen J., Li Z., Chen J., Du Y., Song W., Xuan Z., Zhao L., Song G., Song P., Zheng S. Downregulation of MGMT promotes proliferation of intrahepatic cholangiocarcinoma by regulating p21. *Clin. Transl. Oncol.* 2020;22(3):392-400. DOI 10.1007/s12094-019-02140-9.
- Chen M., Gan X., Yoshino K., Kitakawa M., Shoji I., Deng L., Hotta H. Hepatitis C virus NS5A protein interacts with lysine methyltransferase SET and MYND domain-containing 3 and induces activator protein 1 activation. *Microbiol. Immunol.* 2016;60:407-417. DOI 10.1111/1348-0421.12383.
- Chen P., Meng C., Wang Q., Yang X., Huang Z., Xing X., Lin Y., Liu X., Peng J., Lin Y. Death-associated protein kinase 1 suppresses hepatocellular carcinoma cell migration and invasion by upregulation of DEAD-box helicase 20. *Cancer Sci.* 2020;111(8):2803-2813. DOI 10.1111/cas.14499.
- Cheng J., Wei D., Ji Y., Chen L., Yang L., Li G., Wu L., Hou T., Xie L., Ding G., Li H., Li Y. Integrative analysis of DNA methylation and gene expression reveals hepatocellular carcinoma-specific diagnostic biomarkers. *Genome Med.* 2018;10(1):42. DOI 10.1186/s13073-018-0548-z.
- Choi J., Southworth L.K., Sarin K.Y., Venteicher A.S., Ma W., Chang W., Cheung P., Jun S., Artandi M.K., Shah N., Kim S.K., Artandi S.E. TERT promotes epithelial proliferation through transcriptional control of a Myc- and Wnt-related developmental program. *PLoS Genet.* 2008;4(1):e10. DOI 10.1371/journal.pgen.0040010.
- de Chasse B., Navratil V., Tafforeau L., Hiet M.S., Aublin-Gex A., Agaugué S., Meiffren G., Pradezynski F., Faria B.F., Chantier T., Le Breton M., Pellet J., Davoust N., Mangeot P.E., Chaboud A., Penin F., Jacob Y., Vidalain P.O., Vidal M., André P., Rabourdin-Combe C., Lotteau V. Hepatitis C virus infection protein network. *Mol. Syst. Biol.* 2008;4:230. DOI 10.1038/msb.2008.66.
- Deng Y., Yu B., Cheng Q., Jin J., You H., Ke R., Tang N., Shen Q., Shu H., Yao G., Zhang Z., Qin W. Epigenetic silencing of WIF-1 in hepatocellular carcinomas. *J. Cancer Res. Clin. Oncol.* 2010;136(8):1161-1167. DOI 10.1007/s00432-010-0763-5.
- D'souza S., Lau K.C., Coffin C.S., Patel T.R. Molecular mechanisms of viral hepatitis induced hepatocellular carcinoma. *World J. Gastroenterol.* 2020;26(38):5759-5783. DOI 10.3748/wjg.v26.i38.5759.
- Feng L., Li J., Yan L.D., Tang J. RASSF1A suppresses proliferation of cervical cancer cells. *Asian Pac. J. Cancer Prev.* 2014;15(14):5917-5920. DOI 10.7314/apjcp.2014.15.14.5917.
- Fernández-Barrera M.G., Arechederra M., Colyn L., Berasain C., Avila M.A. Epigenetics in hepatocellular carcinoma development and therapy: the tip of the iceberg. *JHEP Rep.* 2020;2(6):100167. DOI 10.1016/j.jhepr.2020.100167.
- Glotov A.S., Tiys E.S., Vashukova E.S., Pakin V.S., Demenkov P.S., Saik O.V., Ivanisenko T.V., Arzhanova O.N., Mozgovaya E.V., Zainulina M.S., Kolchanov N.A., Baranov V.S., Ivanisenko V.A. Molecular association of pathogenetic contributors to pre-eclampsia (pre-eclampsia associome). *BMC Syst. Biol.* 2015;9(Suppl.2):S4. DOI 10.1186/1752-0509-9-S2-S4.
- Gui Y., Yeganeh M., Ramanathan S., Leblanc C., Pomerleau V., Ferbeyre G., Saucier C., Ilangumaran S. SOCS1 controls liver regeneration by regulating HGF signaling in hepatocytes. *J. Hepatol.* 2011;55(6):1300-1308. DOI 10.1016/j.jhep.2011.03.027.
- Guo N., Chen R., Li Z., Liu Y., Cheng D., Zhou Q., Zhou J., Lin Q. Hepatitis C virus core upregulates the methylation status of the RASSF1A promoter through regulation of SMYD3 in hilar cholangiocarcinoma cells. *Acta Biochim. Biophys. Sin. (Shanghai)*. 2011;43(5):354-361. DOI 10.1093/abbs/gmr021.
- Han Y., Niu J., Wang D., Li Y. Hepatitis C virus protein interaction network analysis based on hepatocellular carcinoma. *PLoS One*. 2016;11(4):e0153882. DOI 10.1371/journal.pone.0153882.
- Hassan M., Ghozlan H., Abdel-Kader O. Activation of c-Jun NH2-terminal kinase (JNK) signaling pathway is essential for the stimulation of hepatitis C virus (HCV) non-structural protein 3 (NS3)-mediated cell growth. *Virology*. 2005;333(2):324-336. DOI 10.1016/j.virol.2005.01.008.
- Hassan M., Selimovic D., Ghozlan H., Abdel-Kader O. Induction of high-molecular-weight (HMW) tumor necrosis factor (TNF) alpha by hepatitis C virus (HCV) non-structural protein 3 (NS3) in liver cells is AP-1 and NF-κB-dependent activation. *Cell. Signal.* 2007;19(2):301-311. DOI 10.1016/j.cellsig.2006.07.002.
- He B., Reguart N., You L., Mazieres J., Xu Z., Lee A.Y., Mikami I., McCormick F., Jablons D.M. Blockade of Wnt-1 signaling induces apoptosis in human colorectal cancer cells containing downstream mutations. *Oncogene*. 2005;24(18):3054-3058. DOI 10.1038/sj.onc.1208511.
- Hernandez-Meza G., von Felden J., Gonzalez-Kozlova E.E., Garcia-Lezana T., Peix J., Portela A., Craig A.J., Sayols S., Schwartz M., Losic B., Mazzaferro V., Esteller M., Llovet J.M., Villanueva A. DNA methylation profiling of human hepatocarcinogenesis. *Hepatology*. 2021;74(1):183-199. DOI 10.1002/hep.31659.
- Hoshida Y., Villanueva A., Kobayashi M., Peix J., Chiang D.Y., Camargo A., Gupta S., Moore J., Wrobel M.J., Lerner J., Reich M., Chan J.A., Glickman J.N., Ikeda K., Hashimoto M., Watanabe G., Daidone M.G., Roayaie S., Schwartz M., Thung S., Salvesen H.B., Gabriel S., Mazzaferro V., Bruix J., Friedman S.L., Kumada H., Llovet J.M., Golub T.R. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N. Engl. J. Med.* 2008;359(19):1995-2004. DOI 10.1056/NEJMoa0804525.
- Huang L., Li M.X., Wang L., Li B.K., Chen G.H., He L.R., Xu L., Yuan Y.F. Prognostic value of Wnt inhibitory factor-1 expression in hepatocellular carcinoma that is independent of gene methylation. *Tumour Biol.* 2011;32(1):233-240. DOI 10.1007/s13277-010-0117-6.
- In der Stroth L., Tharehalli U., Günes C., Lechel A. Telomeres and telomerase in the development of liver cancer. *Cancers (Basel)*. 2020;12(8):2048. DOI 10.3390/cancers12082048.
- Ivanisenko N.V., Seyrek K., Kolchanov N.A., Ivanisenko V.A., Lavrik I.N. The role of death domain proteins in host response upon SARS-CoV-2 infection: modulation of programmed cell death and translational applications. *Cell Death Discov.* 2020;6:101. DOI 10.1038/s41420-020-00331-w.
- Ivanisenko T.V., Saik O.V., Demenkov P.S., Ivanisenko N.V., Savostianov A.N., Ivanisenko V.A. ANDDigest: a new web-based module of ANDSystem for the search of knowledge in the scientific literature. *BMC Bioinformatics*. 2020;21(Suppl.11):228. DOI 10.1186/s12859-020-03557-8.
- Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics*. 2019;20(Suppl.1):34. DOI 10.1186/s12859-018-2567-6.
- Ivanisenko V.A., Gaisler E.V., Basov N.V., Rogachev A.D., Cherezis S.V., Ivanisenko T.V., Demenkov P.S., Mishchenko E.L., Khrip-

- ko O.P., Khripko Yu.I., Voevoda S.M., Karpenko T.N., Velichko A.J., Voevoda M.I., Kolchanov N.A., Pokrovsky A.G. Plasma metabolomics and gene regulatory networks analysis reveal the role of non-structural SARS-CoV-2 viral proteins in metabolic dysregulation in COVID-19 patients. *Sci. Rep.* 2022;12:19977. DOI 10.1038/s41598-022-24170-0.
- Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst. Biol.* 2015;9(Suppl.2):S2. DOI 10.1186/1752-0509-9-S2-S2.
- Iwai A., Takegami T., Shiozaki T., Miyazaki T. Hepatitis C virus NS3 protein can activate the Notch-signaling pathway through binding to a transcription factor, SRCAP. *PLoS One.* 2011;6(6):e20718. DOI 10.1371/journal.pone.0020718.
- Jardin F., Ruminy P., Bastard C., Tilly H. The BCL6 proto-oncogene: a leading role during germinal center development and lymphomagenesis. *Pathol. Biol. (Paris).* 2007;55(1):73-83. DOI 10.1016/j.patbio.2006.04.001.
- Jaroszewicz J., Flisiak-Jackiewicz M., Lebensztejn D., Flisiak R. Current drugs in early development for treating hepatitis C virus-related hepatic fibrosis. *Expert Opin. Investig. Drugs.* 2015;24(9):1229-1239. DOI 10.1517/13543784.2015.1057568.
- Jiang L.H., Hao Y.L., Zhu J.W. Expression and prognostic value of HER-2/neu, STAT3 and SOCS3 in hepatocellular carcinoma. *Clin. Res. Hepatol. Gastroenterol.* 2019;43(3):282-291. DOI 10.1016/j.clinre.2018.09.011.
- Jing W., Peng R., Li X., Lv S., Duan Y., Jiang S. Study on the prognostic values of TTC36 correlated with immune infiltrates and its methylation in hepatocellular carcinoma. *J. Immunol. Res.* 2022;2022:7267131. DOI 10.1155/2022/7267131.
- Kim B.R., Park S.H., Jeong Y.A., Na Y.J., Kim J.L., Jo M.J., Jeong S., Yun H.K., Oh S.C., Lee D.H. RUNX3 enhances TRAIL-induced apoptosis by upregulating DR5 in colorectal cancer. *Oncogene.* 2019;38:3903-3918. DOI 10.1038/s41388-019-0693-x.
- Kohsaka S., Wang L., Yachi K., Mahabir R., Narita T., Itoh T., Tani-no M., Kimura T., Nishihara H., Tanaka S. STAT3 inhibition overcomes temozolomide resistance in glioblastoma by downregulating MGMT expression. *Mol. Cancer Ther.* 2012;11(6):1289-1299. DOI 10.1158/1535-7163.MCT-11-0801.
- Konnikova L., Simeone M.C., Kruger M.M., Kotecki M., Cochran B.H. Signal transducer and activator of transcription 3 (STAT3) regulates human telomerase reverse transcriptase (hTERT) expression in human cancer and primary cells. *Cancer Res.* 2005;65(15):6516-6520. DOI 10.1158/0008-5472.CAN-05-0924.
- Li B., Li X., Li Y., Guo H., Sun S.Y., He Q.Q., Wang Y., Luo J., Wen J.F., Zheng H., Feng D.Y. The effects of hepatitis C virus non-structural protein 3 on cell growth mediated by extracellular signal-related kinase cascades in human hepatocytes in vitro. *Int. J. Mol. Med.* 2010;26(2):273-279. DOI 10.3892/ijmm\_00000462.
- Lin B., Hong H., Jiang X., Li C., Zhu S., Tang N., Wang X., She F., Chen Y. c-Jun suppresses the expression of WNT inhibitory factor 1 through transcriptional regulation and interaction with DNA methyltransferase 1 in gallbladder cancer. *Mol. Med. Rep.* 2018;17(6):8180-8188. DOI 10.3892/mmr.2018.8890.
- Liu J., Ma Q., Zhang M., Wang X., Zhang D., Li W., Wang F., Wu E. Alterations of TP53 are associated with a poor out-come for patients with hepatocellular carcinoma: evidence from a systematic review and meta-analysis. *Eur. J. Cancer.* 2012;48(15):2328-2338. DOI 10.1016/j.ejca.2012.03.001.
- Llovet J.M., Zucman-Rossi J., Pikarsky E., Sangro B., Schwartz M., Sherman M., Gores G. Hepatocellular carcinoma. *Nat. Rev. Dis. Primers.* 2016;2:16018. DOI 10.1038/nrdp.2016.18.
- Loeb D.M. WT1 influences apoptosis through transcriptional regulation of Bcl-2 family members. *Cell Cycle.* 2006;5(12):1249-1253. DOI 10.4161/cc.5.12.2807.
- Machida K., Cheng K.T., Lai C.K., Jeng K.S., Sung V.M., Lai M.M. Hepatitis C virus triggers mitochondrial permeability transition with production of reactive oxygen species, leading to DNA damage and STAT3 activation. *J. Virol.* 2006;80(14):7199-7207. DOI 10.1128/JVI.00321-06.
- Machida K., Cheng K.T., Sung V.M., Shimodaira S., Lindsay K.L., Levine A.M., Lai M.Y., Lai M.M. Hepatitis C virus induces a mutator phenotype: enhanced mutations of immunoglobulin and proto-oncogenes. *Proc. Natl. Acad. Sci. USA.* 2004;101(12):4262-4267. DOI 10.1073/pnas.0303971101.
- Mayo M.W., Wang C.Y., Drouin S.S., Madrid L.V., Marshall A.F., Reed J.C., Weissman B.E., Baldwin A.S. WT1 modulates apoptosis by transcriptionally upregulating the *bcl-2* proto-oncogene. *EMBO J.* 1999;18(14):3990-4003. DOI 10.1093/emboj/18.14.3990.
- McGlynn K.A., Petrick J.L., El-Serag H.B. Epidemiology of hepatocellular carcinoma. *Hepatology.* 2021;73(Suppl.1):4-13. DOI 10.1002/hep.31288.
- Mžik M., Chmelařová M., John S., Laco J., Slabý O., Kiss I., Bohovicová L., Palička V., Nekvindová J. Aberrant methylation of tumour suppressor genes *WT1*, *GATA5* and *PAX5* in hepatocellular carcinoma. *Clin. Chem. Lab. Med.* 2016;54(12):1971-1980. DOI 10.1515/cclm-2015-1198.
- Nault J.C., Ningharhari M., Rebouissou S., Zucman-Rossi J. The role of telomeres and telomerase in cirrhosis and liver cancer. *Nat. Rev. Gastroenterol. Hepatol.* 2019;16(9):544-558. DOI 10.1038/s41575-019-0165-3.
- Neumann O., Kesselmeier M., Geffers R., Pellegrino R., Radlwimmer B., Hoffmann K., Ehemann V., Schemmer P., Schirmacher P., Lorenzo Bermejo J., Longerich T. Methylation analysis and integrative profiling of human HCCs identify novel protumorigenic factors. *Hepatology.* 2012;56(5):1817-1827. DOI 10.1002/hep.25870.
- Ni Y., Gu J., Wu J., Xu L., Rui Y. MGMT-mediated neuron apoptosis in injured rat spinal cord. *Tissue Cell.* 2020;62:101311. DOI 10.1016/j.tice.2019.101311.
- Nowyhed H.N., Huynh T.R., Blatchley A., Wu R., Thomas G.D., Hedrick C.C. The nuclear receptor Nr4a1 controls CD8 T cell development through transcriptional suppression of Runx3. *Sci. Rep.* 2015;5:9059. DOI 10.1038/srep09059.
- Oversoe S.K., Clement M.S., Pedersen M.H., Weber B., Aagaard N.K., Villadsen G.E., Grønbaek H., Hamilton-Dutoit S.J., Sorensen B.S., Kelsen J. TERT promoter mutated circulating tumor DNA as a biomarker for prognosis in hepatocellular carcinoma. *Scand. J. Gastroenterol.* 2020;55(12):1433-1440. DOI 10.1080/00365521.2020.1837928.
- Papic N., Maxwell C.I., Delker D.A., Liu S., Heale B.S., Hagedorn C.H. RNA-sequencing analysis of 5' capped RNAs identifies many new differentially expressed genes in acute hepatitis C virus infection. *Viruses.* 2012;4:581-612. DOI 10.3390/v4040581.
- Phan R.T., Dalla-Favera R. The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells. *Nature.* 2004;432(7017):635-639. DOI 10.1038/nature03147.
- Rabaan A.A., Al-Ahmed S.H., Bazzi A.M., Alfouzan W.A., Alsuliman S.A., Aldrazi F.A., Haque S. Overview of hepatitis C infection, molecular biology, and new treatment. *J. Infect. Public Health.* 2020;13(5):773-783. DOI 10.1016/j.jiph.2019.11.015.
- Reville K., Wang T., Lachenmayer A., Kojima K., Harrington A., Li J., Hoshida Y., Llovet J.M., Powers S. Genome-wide methylation analysis and epigenetic unmasking identify tumor suppressor genes in hepatocellular carcinoma. *Gastroenterology.* 2013;145(6):1424-1435.e1-25. DOI 10.1053/j.gastro.2013.08.055.
- Saik O.V., Demenkov P.S., Ivanisenko T.V., Bragina E.Y., Freidin M.B., Goncharova I.A., Dosenko V.E., Zolotareva O.I., Hofstaedt R., Lavrik I.N., Rogaev E.I., Ivanisenko V.A. Novel candidate genes important for asthma and hypertension comorbidity revealed from associative gene networks. *BMC Med. Genomics.* 2018;11(Suppl.1):15. DOI 10.1186/s12920-018-0331-4.
- Saik O.V., Ivanisenko T.V., Demenkov P.S., Ivanisenko V.A. Interaction of the hepatitis C virus: literature mining with ANDSystem. *Virus Res.* 2016;218:40-48. DOI 10.1016/j.virusres.2015.12.003.

- Sarin K.Y., Cheung P., Gilson D., Lee E., Tennen R.I., Wang E., Artandi M.K., Oro A.E., Artandi S.E. Conditional telomerase induction causes proliferation of hair follicle stem cells. *Nature*. 2005; 436(7053):1048-1052. DOI 10.1038/nature03836.
- Sawangarun W., Mandasari M., Aida J., Morita K.I., Kayamori K., Ikeda T., Sakamoto K. Loss of Notch1 predisposes oro-esophageal epithelium to tumorigenesis. *Exp. Cell Res.* 2018;372(2):129-140. DOI 10.1016/j.yexcr.2018.09.019.
- Schulze K., Imbeaud S., Letouzé E., Alexandrov L.B., Calderaro J., Rebouissou S., Couchy G., Meiller C., Shinde J., Soysouvanh F., Calatayud A.L., Pinyol R., Pelletier L., Balabaud C., Laurent A., Blanc J.F., Mazzaferro V., Calvo F., Villanueva A., Nault J.C., Bioulac-Sage P., Stratton M.R., Llovet J.M., Zucman-Rossi J. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat. Genet.* 2015;47(5): 505-511. DOI 10.1038/ng.3252.
- Sera T., Hiasa Y., Mashiba T., Tokumoto Y., Hirooka M., Konishi I., Matsuura B., Michitaka K., Udaka K., Onji M. Wilms' tumour 1 gene expression is increased in hepatocellular carcinoma and associated with poor prognosis. *Eur. J. Cancer*. 2008;44(4):600-608. DOI 10.1016/j.ejca.2008.01.008.
- Song Z., Li Z., Han W., Zhu C., Lou N., Li X., Luo G., Peng S., Li G., Zhao Y., Guo Y. Low DAPK1 expression correlates with poor prognosis and sunitinib resistance in clear cell renal cell carcinoma. *Aging (Albany NY)*. 2020;13(2):1842-1858. DOI 10.18632/aging.103638.
- Takakura M., Kyo S., Inoue M., Wright W.E., Shay J.W. Function of AP-1 in transcription of the telomerase reverse transcriptase gene (*TERT*) in human and mouse cells. *Mol. Cell Biol.* 2005;25(18): 8037-8043. DOI 10.1128/MCB.25.18.8037-8043.2005.
- Tan Y., Li Y. HCV core protein promotes hepatocyte proliferation and chemoresistance by inhibiting NR4A1. *Biochem. Biophys. Res. Commun.* 2015;466(3):592-598. DOI 10.1016/j.bbrc.2015.09.091.
- Tucci F.A., Broering R., Johansson P., Schlaak J.F., Küppers R. B cells in chronically hepatitis C virus-infected individuals lack a virus-induced mutation signature in the *TP53*, *CTNNB1*, and *BCL6* genes. *J. Virol.* 2013;87(5):2956-2962. DOI 10.1128/JVI.03081-12.
- Xu R.H., Wei W., Krawczyk M., Wang W., Luo H., Flagg K., Yi S., Shi W., Quan Q., Li K., Zheng L., Zhang H., Caughey B.A., Zhao Q., Hou J., Zhang R., Xu Y., Cai H., Li G., Hou R., Zhong Z., Lin D., Fu X., Zhu J., Duan Y., Yu M., Ying B., Zhang W., Wang J., Zhang E., Zhang C., Li O., Guo R., Carter H., Zhu J.K., Hao X., Zhang K. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat. Mater.* 2017; 16(11):1155-1161. DOI 10.1038/nmat4997.
- Yang C., Zhang Y., Wang J., Li L., Wang L., Hu M., Xu M., Long Y., Rong R., Zhu T. A novel cyclic helix B peptide inhibits dendritic cell maturation during amelioration of acute kidney graft rejection through Jak-2/STAT3/SOCS1. *Cell Death Dis.* 2015;6(11):e1993. DOI 10.1038/cddis.2015.338.
- Ye S., Zhao X.Y., Hu X.G., Li T., Xu Q.R., Yang H.M., Huang D.S., Yang L. TP53 and RET may serve as biomarkers of prognostic evaluation and targeted therapy in hepatocellular carcinoma. *Oncol. Rep.* 2017;37(4):2215-2226. DOI 10.3892/or.2017.5494.
- Zhang C., Li J., Huang T., Duan S., Dai D., Jiang D., Sui X., Li D., Chen Y., Ding F., Huang C., Chen G., Wang K. Meta-analysis of DNA methylation biomarkers in hepatocellular carcinoma. *Oncotarget*. 2016;7(49):81255-81267. DOI 10.18632/oncotarget.13221.
- Zhang H., Weng X., Ye J., He L., Zhou D., Liu Y. Promoter hypermethylation of *TERT* is associated with hepatocellular carcinoma in the Han Chinese population. *Clin. Res. Hepatol. Gastroenterol.* 2015;39(5):600-609. DOI 10.1016/j.clinre.2015.01.002.
- Zhu Z., Tran H., Mathahs M.M., Moninger T.O., Schmidt W.N. HCV induces telomerase reverse transcriptase, increases its catalytic activity, and promotes caspase degradation in infected human hepatocytes. *PLoS One*. 2017;12(1):e0166853. DOI 10.1371/journal.pone.0166853.
- Zhu Z., Wilson A.T., Gopalakrishna K., Brown K.E., Luxon B.A., Schmidt W.N. Hepatitis C virus core protein enhances Telomerase activity in Huh7 cells. *J. Med. Virol.* 2010;82(2):239-248. DOI 10.1002/jmv.21644.
- Zong C., Qin D., Yu C., Gao P., Chen J., Lu S., Zhang Y., Liu Y., Yang Y., Pu Z., Li X., Fu Y., Guan Q., Wang X. The stress-response molecule NR4A1 resists ROS-induced pancreatic  $\beta$ -cells apoptosis via WT1. *Cell Signal*. 2017;35:129-139. DOI 10.1016/j.cellsig.2017.03.012.
- Zucman-Rossi J., Villanueva A., Nault J.C., Llovet J.M. Genetic landscape and biomarkers of hepatocellular carcinoma. *Gastroenterology*. 2015;149(5):1226-1239.e4. DOI 10.1053/j.gastro.2015.05.061.

#### ORCID ID

E.A. Antropova orcid.org/0000-0003-2158-3252  
P.S. Demenkov orcid.org/0000-0001-9433-8341

**Благодарности.** Исследование выполнено при финансовой поддержке проекта № 075-15-2021-944 Министерства науки и высшего образования РФ в рамках ERA-NET «Идентификация мишеней и разработка лекарственных средств при раке печени (TAIGA)».

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 15.11.2022. После доработки 22.11.2022. Принята к публикации 22.11.2022.

Перевод на английский язык <https://vavilov.elpub.ru/jour>

## Рациональная метаболическая инженерия *Corynebacterium glutamicum* для продукции L-валина

М.Е. Шереметьева<sup>1</sup>✉, К.Э. Ануфриев<sup>1</sup>, Т.М. Хлебодарова<sup>2,3</sup>, Н.А. Колчанов<sup>2,3</sup>, А.С. Яненко<sup>1</sup>

<sup>1</sup> Национальный исследовательский центр «Курчатовский институт», Курчатовский геномный центр, Москва, Россия

<sup>2</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>3</sup> Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия

✉ m.e.sheremetieva@gmail.com

**Аннотация.** L-Валин – одна из девяти аминокислот, которые не могут быть синтезированы *de novo* высшими организмами и должны поступать с пищей. Эта аминокислота не только служит строительным материалом для белков, но также регулирует белковый и энергетический обмен и участвует в нейротрансмиссии. L-Валин используется в пищевой и фармацевтической промышленности, медицине и косметике, но в первую очередь в качестве кормовой добавки для животных. Добавление L-валина в корм отдельно или в смеси с другими незаменимыми аминокислотами позволяет использовать корма с меньшим содержанием сырого белка, повышает качество и количество мяса свиней и цыплят-бройлеров, а также улучшает репродуктивные функции сельскохозяйственных животных. Несмотря на то что рынок L-валина постоянно растет, в нашей стране эта аминокислота пока не производится. В современных условиях создание штаммов-продуцентов и организация производства L-валина для России особенно актуальны. Один из наиболее часто используемых базовых микроорганизмов для создания продуцентов аминокислот наряду с *Escherichia coli* – почвенная бактерия *Corynebacterium glutamicum*. Обзор посвящен анализу основных стратегий разработки продуцентов L-валина на базе *C. glutamicum*. Рассмотрены различные аспекты биосинтеза L-валина у коринебактерий: биохимия, стехиометрия и регуляция процесса, ферменты и соответствующие им гены, системы экспорта и импорта, связь биосинтеза L-валина с центральным метаболизмом клетки. Выявлены ключевые генетические элементы для создания штаммов-продуцентов на основе *C. glutamicum*. Описано использование метаболической инженерии для усиления реакций биосинтеза L-валина и уменьшения образования побочных продуктов. Показаны перспективы усовершенствования штаммов с точки зрения повышения их продуктивности и улучшения технологических характеристик. Информация, представленная в обзоре, может быть использована при получении продуцентов других аминокислот с разветвленной боковой цепью – L-лейцина и L-изолейцина, а также D-пантотената.

Ключевые слова: *Corynebacterium glutamicum*; L-валин; метаболическая инженерия; штамм-продуцент.

**Для цитирования:** Шереметьева М.Е., Ануфриев К.Э., Хлебодарова Т.М., Колчанов Н.А., Яненко А.С. Рациональная метаболическая инженерия *Corynebacterium glutamicum* для продукции L-валина. *Вавиловский журнал генетики и селекции*. 2022;26(8):743-757. DOI 10.18699/VJGB-22-90

## Rational metabolic engineering of *Corynebacterium glutamicum* to create a producer of L-valine

M.E. Sheremetieva<sup>1</sup>✉, K.E. Anufriev<sup>1</sup>, T.M. Khlebodarova<sup>2,3</sup>, N.A. Kolchanov<sup>2,3</sup>, A.S. Yanenko<sup>1</sup>

<sup>1</sup> NRC “Kurchatov Institute”, Kurchatov Genomic Center, Moscow, Russia

<sup>2</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>3</sup> Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia

✉ m.e.sheremetieva@gmail.com

**Abstract.** L-Valine is one of the nine amino acids that cannot be synthesized *de novo* by higher organisms and must come from food. This amino acid not only serves as a building block for proteins, but also regulates protein and energy metabolism and participates in neurotransmission. L-Valine is used in the food and pharmaceutical industries, medicine and cosmetics, but primarily as an animal feed additive. Adding L-valine to feed, alone or mixed with other essential amino acids, allows for feeds with lower crude protein content, increases the quality and quantity of pig meat and broiler chicken meat, as well as improves reproductive functions of farm animals. Despite the fact that the market for L-valine is constantly growing, this amino acid is not yet produced in our country. In modern conditions, the creation of strains-producers and organization of L-valine production are especially relevant for Russia. One of the basic microorganisms most commonly used for the creation of amino acid producers, along with *Escherichia coli*, is the soil bacterium *Corynebacterium glutamicum*. This review is devoted to the analysis of the main strategies for the development of L-valine producers based on *C. glutamicum*. Various aspects of L-valine biosynthesis in *C. glutamicum* are reviewed: process biochemistry, stoichiometry and regulation, enzymes and their corresponding genes, export and import systems, and the relationship of L-valine biosynthesis with central cell metabolism. Key genetic elements for the creation

of *C. glutamicum*-based strains-producers are identified. The use of metabolic engineering to enhance L-valine biosynthesis reactions and to reduce the formation of byproducts is described. The prospects for improving strains in terms of their productivity and technological characteristics are shown. The information presented in the review can be used in the production of producers of other amino acids with a branched side chain, namely L-leucine and L-isoleucine, as well as D-pantothenate.

**Key words:** *Corynebacterium glutamicum*; L-valine; metabolic engineering; producer strain.

**For citation:** Sheremetieva M.E., Anufriev K.E., Khlebodarova T.M., Kolchanov N.A., Yanenko A.S. Rational metabolic engineering of *Corynebacterium glutamicum* to create a producer of L-valine. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):743-757. DOI 10.18699/VJGB-22-90

## Введение

L-Валин относится к группе протеиногенных аминокислот с разветвленной боковой цепью (branched-chain amino acids, ВСАА), куда также входят L-лейцин и L-изолейцин (далее – валин, лейцин, изолейцин). Это незаменимые аминокислоты, которые не синтезируются в организме человека и животных и должны присутствовать в рационе. Поэтому такие аминокислоты используются главным образом в индустрии кормов для животных и в качестве пищевой добавки для человека (Karau, Grayson, 2014). Первое особенно актуально для решения общемировой задачи интенсификации животноводства. Добавление к кормам валина, как отдельно, так и в смеси с другими ВСАА, приводит к повышению качества и количества мяса свиней и цыплят-бройлеров, повышает яйценоскость кур, увеличивает лактацию, жирность молока и аппетит у свиней (Zheng et al., 2017; Che et al., 2021; Jian et al., 2021). Необходимо, однако, соблюдать баланс между различными ВСАА, нарушение которого может снижать наблюдаемые положительные эффекты (Holen et al., 2022).

Помимо животноводства и пищевой промышленности, ВСАА находят свое применение в фармакологии и медицине. ВСАА не только служат строительным материалом для белков, но также участвуют в регуляции белкового и энергетического обмена, их потребление повышает толерантность к физическим нагрузкам и ускоряет окисление жирных кислот (Kainulainen et al., 2013). Они полезны как биологические добавки при хронических заболеваниях печени (Kawaguchi et al., 2011) и для стимуляции макрофагального фагоцитоза бактериальных патогенов со множественной лекарственной устойчивостью (Chen et al., 2017). Как и в случае с кормовыми добавками, при использовании ВСАА для производства продуктов питания и лекарств следует внимательно относиться к подбору концентраций. Избыток ВСАА в плазме крови человека повышает риск развития ряда заболеваний, включая диабет второго типа, метаболический синдром, ожирение, гипертонию и сердечно-сосудистые заболевания (Holeček, 2018; Dimou et al., 2022), но практически не сказывается на спортсменах, для которых характерны высокие физические нагрузки (Shou et al., 2019).

На мировом рынке кормовых добавок, объем которого в 2022 г., по прогнозам, может составить 34.2 млрд долларов США, аминокислоты занимают 62.3 %. Наиболее востребованы L-лизин и L-метионин (далее – лизин, метионин); рынок валина – один из самых быстрорастущих, наряду с L-треонином (далее – треонин) и L-триптофаном. Потребление кормовых аминокислот сосредоточено в Европе, США и Китае, доля России составляет ме-

нее 2 %, однако проявляет тенденцию к росту – с 2016 по 2017 г. прибавка составила 2.9 % (<https://agri-news.ru/zhurnal/2018/32018/ekonomika-menedzhment-ryinki/ryinok-kormovyix-aminokislota.html>). В настоящее время весь валин на российском рынке импортируется из Китая, одного из основных производителей этой аминокислоты.

Аминокислоты можно выделять из природных источников белка, получать путем химического синтеза, а также микробиологическим способом, основанным на применении штаммов-продуцентов. Последний вариант обладает важными достоинствами: он позволяет использовать возобновляемые сырьевые ресурсы и получать биологически активные L-энантиомеры аминокислот отдельно, а не в смеси с D-энантиомерами, и поэтому применяется ведущими производителями валина (D'Este et al., 2017).

Продуценты аминокислот разрабатывают на базе *Escherichia coli* и *Corynebacterium glutamicum*. *E. coli* – досконально изученная бактерия, для которой имеется обширный инструментарий генетической модификации, в связи с чем ранее штаммы-продуценты получали преимущественно на ее основе. Однако использовались и штаммы *C. glutamicum*, созданные с помощью селекции, история их применения для производства аминокислот насчитывает уже более 60 лет (Leuchtenberger et al., 2005). В последние десятилетия, достигнув значительного прогресса как в понимании метаболизма *C. glutamicum*, так и в усовершенствовании методов модификации их генома, разработчики штаммов-продуцентов все чаще стали отдавать предпочтение коринебактериям.

Это непатогенные, GC-богатые грамположительные бактерии, которые, в отличие от *E. coli*, не образуют эндотоксинов, вызывающих аллергические реакции у высших организмов, а кроме того, характеризуются гибким клеточным метаболизмом, генетической стабильностью, стрессоустойчивостью, включая устойчивость к высоким концентрациям источников углерода и метаболитов, и способностью синтезировать целевой продукт при остановке роста (Varitugo et al., 2018). В настоящее время валин, полученный путем ферментации с использованием штаммов *C. glutamicum*, признан безопасным (нетоксичным и неканцерогенным) для применения в качестве пищевой и кормовой добавки и в других биологических целях (Kang et al., 2020).

В обзоре представлены основные стратегии повышения продуктивности валина клетками *C. glutamicum* и суммированы достижения по созданию штаммов-продуцентов валина. Помимо получения валина, рассматриваются некоторые аспекты получения изолейцина, лейцина и D-пантотена-

та (далее – пантотенат), поскольку в биосинтезе данных соединений участвуют общие с валином метаболические предшественники, кофакторы и ферменты.

### Биосинтез валина у *C. glutamicum* и механизмы регуляции данного процесса

Валин (2-амино-3-метилбутановая кислота) синтезируется из пирувата (пировиноградной кислоты) в результате четырех последовательных реакций, включающих (рис. 1): 1) конденсацию двух молекул пирувата с образованием 2-ацетолактата, катализируемую ацетолактатсинтазой (ANAS); 2) НАДФН-зависимое превращение 2-ацетолактата в 2,3-диоксиизовалерат, катализируемое ацетолактатредуктоизомеразой (ANAIR); 3) превращение 2,3-диоксиизовалерата в 2-оксоизовалерат, катализируемое дегидратазой дигидроксикислот (DHAD); 4) НАДФН-зависимое образование валина из 2-оксоизовалерата, катализируемое трансаминазой ВСАА (BCAT) и рядом других трансаминаз (Yamamoto et al., 2017).

В ходе синтеза на продукцию 1 моля валина расходуются 2 моля пирувата и 2 моля восстановительных эквивалентов в форме восстановленного никотинамиддинуклеотидфосфата (НАДФН). Пируват образуется из фосфоенолпирувата (ФЕП) в гликолизе, в ходе которого 1 моль глюкозы превращается в 2 моля пирувата. Основной источник НАДФН у коринебактерий – пентозофосфатный восстановительный путь (ПФП) (Marx et al., 1997).

2-Оксоизовалерат является также предшественником биосинтеза лейцина и пантотената (Park, Lee, 2010). У большинства микроорганизмов, включая *C. glutamicum*, эти же четыре фермента катализируют биосинтез изолейцина из пирувата и 2-оксобутирата. Последний образуется из треонина с помощью треониндегидратазы (TD). Таким образом, процессы биосинтеза всех трех ВСАА (валина, лейцина и изолейцина) тесно связаны. Синтезированные ВСАА выводятся из клетки с помощью одной системы экспорта – BrnFE (Lange et al., 2012).

Схема биосинтеза валина и связанных с этим процессом метаболических путей у *C. glutamicum* представлена на рис. 1. Ключевой фермент путей биосинтеза валина и других ВСАА – ацетолактатсинтаза ANAS, которая катализирует образование либо 2-ацетолактата из двух молекул пирувата (при биосинтезе валина и лейцина), либо образование 2-ацето,2-оксибутирата из пирувата и 2-оксобутирата (при биосинтезе изолейцина). В отличие от *E. coli*, у *C. glutamicum* обнаружена только одна форма фермента ANAS (Keilhauer et al., 1993), тетрамер, состоящий из двух каталитических и двух регуляторных субъединиц (Liu et al., 2016). Каталитическая и регуляторная субъединицы ANAS кодируются генами *ilvB* и *ilvN* соответственно. Вместе с геном *ilvC*, кодирующим ацетолактатредуктоизомеразу ANAIR, эти два гена образуют оперон *ilvBNC* с двумя дополнительными промоторами внутри него. Экспрессия с трех промоторов ведет к образованию транскриптов различной длины (рис. 2). Ген *ilvC* транскрибируется в составе всех мРНК, эффективность его экспрессии самая высокая среди трех генов (Keilhauer et al., 1993; Morbach et al., 2000).

Экспрессия оперона *ilvBNC*, как считается, контролируется через механизм аттенуации транскрипции (англ.

transcription attenuation – транскрипционное ослабление), реализуемый за счет формирования на транскрипте вторичных РНК-структур (шпилек) – терминаторов транскрипции, возникающих в регуляторной области в присутствии высоких концентраций ВСАА (см. рис. 2). Регуляторная область расположена перед геном *ilvB*; помимо участков, ответственных за формирование шпилек, она кодирует также лидерный пептид (25 аминокислот), обогащенный остатками изолейцина (2), валина (3) и лейцина (2). Предполагается, что этот пептид является сенсорным элементом регуляторной системы – при низкой концентрации какой-либо из ВСАА в клетках его трансляция замедляется, в результате чего не происходит образования терминаторных шпилек.

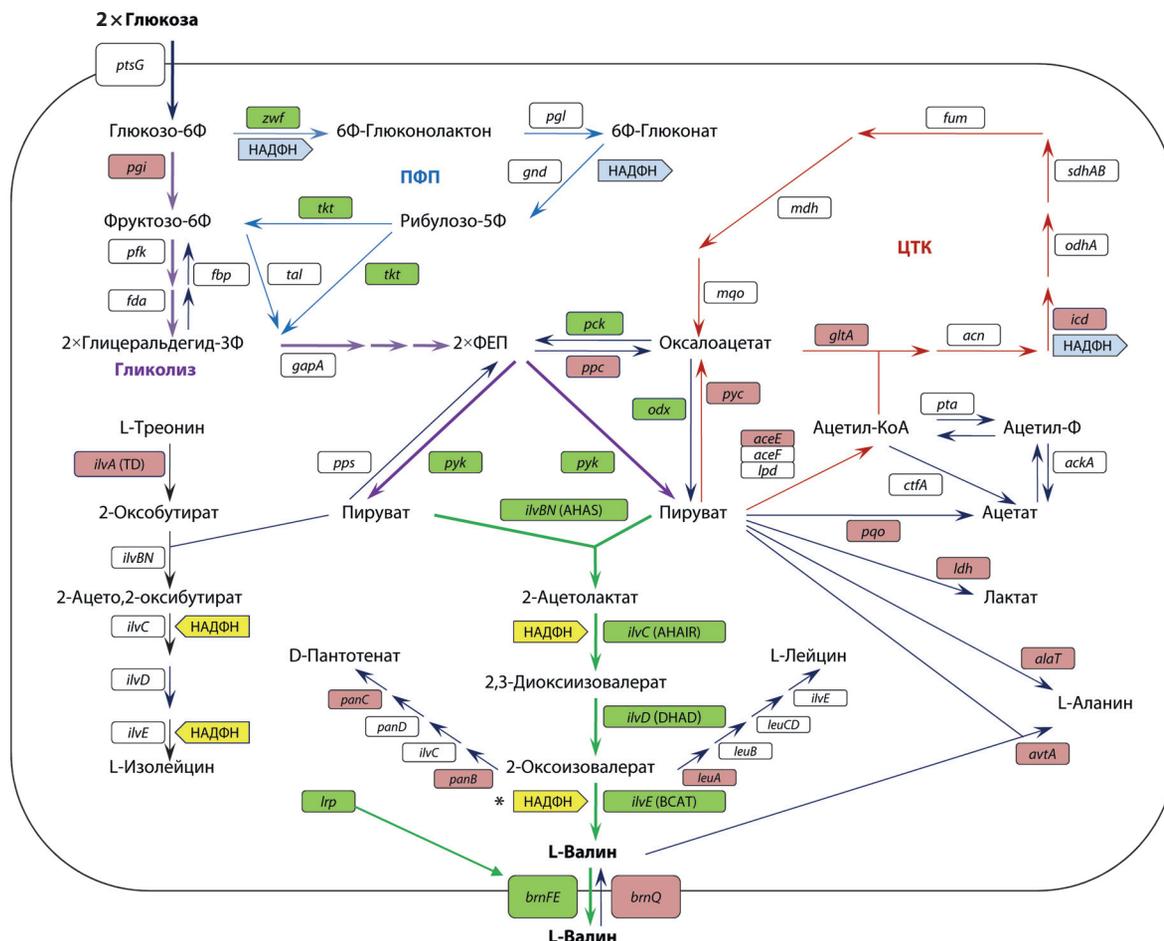
При недостатке одной или нескольких ВСАА экспрессия оперона *ilvBNC* увеличивалась вдвое; замена в лидерном пептиде остатков валина на остатки аланина вела к потере влияния валина на экспрессию (Morbach et al., 2000). Отмечено существенное усиление экспрессии оперона *ilvBNC* в присутствии 2-оксобутирата (Eggeling et al., 1987; Keilhauer et al., 1993; Morbach et al., 2000). Механизм такой регуляции не исследован.

Активность ANAS строго ингибируется валином ( $K_i = 0.9$  мМ), а также лейцином ( $K_i = 6.0$  мМ) и изолейцином ( $K_i = 3.1$  мМ) по механизму обратной связи через присоединение аминокислот к регуляторной субъединице фермента (Eggeling et al., 1987; Morbach et al., 2000; Leyval et al., 2003; Elišáková et al., 2005), а также конкурентно ингибируется 2-оксоизовалератом (Krause et al., 2010a). Независимо от количества присутствующих ВСАА (одной, двух или всех трех), степень ингибирования активности ANAS не превышает 57 % (Elišáková et al., 2005).

Необходимо отметить, что ANAS имеет более низкую субстратную специфичность по отношению к пирувату ( $K_m = 8.3$  мМ) (Leyval et al., 2003), чем к 2-оксобутирату ( $K_m = 4.8$  мМ) (Eggeling et al., 1987), поэтому при прочих равных условиях предпочтение отдается реакции конденсации пирувата с 2-оксобутиратом, ведущей к синтезу изолейцина.

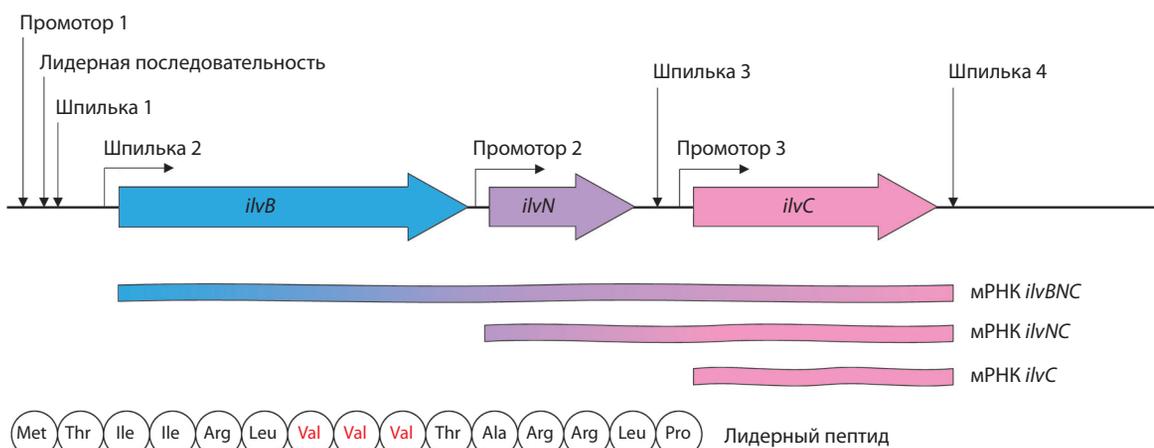
Что касается ANAIR (продукт гена *ilvC*), катализирующей этап изомеризации и превращения 2-ацетолактата в 2,3-диоксиизовалерат, а также 2-ацето,2-оксибутирата в 2,3-диокси-3-метилвалерат в пути синтеза изолейцина, то ее активность зависит от присутствия НАДФН и ингибируется по механизму обратной связи валином и лейцином, но не изолейцином (Leyval et al., 2003; Lee et al., 2019).

О регуляции активности ферментов, контролируемых третьим и четвертым, заключительным, этапами синтеза валина у *C. glutamicum*, а также о регуляции генов, кодирующих их структуру, информации мало. Известно только, что активность дегидратазы дигидроксикислот DHAD (продукт гена *ilvD*) слабо ингибируется валином и лейцином и не ингибируется изолейцином (Leyval et al., 2003), а активность трансаминазы BCAT (продукт гена *ilvE*) зависит от доступности НАДФН. Донором аминокислоты в реакции трансаминирования служит L-глутамат (далее – глутамат), который превращается в 2-оксоглутаровую кислоту; НАДФН необходим для регенерации глутамата с помощью глутаматдегидрогеназы. По-



**Рис. 1.** Биосинтез валина и связанные с данным процессом метаболические пути в клетках *C. glutamicum*.

Выделены гены, усиление экспрессии которых приводит к повышению (зеленый цвет) или снижению (красный цвет) продукции валина. Подробное описание и расшифровка сокращений приведены в тексте. Звездочкой обозначена реакция, в которой НАДФН используется опосредованно (см. пояснения в тексте).



**Рис. 2.** Организация оперона *ilvBNC* у *C. glutamicum* и регуляция его экспрессии (адаптировано из обзора (Wang et al., 2018)).

казано, что реакция, катализируемая глутаматдегидрогеназой, является основной реакцией ассимиляции азота в условиях избытка аммиака, которые обычно имеют место в процессах производства аминокислот (Burkovski, 2003).

Показано также участие в биосинтезе валина аланин/валинтрансминазы (продукт гена *avtA*), которая использует в качестве донора аминогруппы не глутамат, а L-аланин (далее – аланин) или α-аминобутират (Leyval et al., 2003).

Анализ динамики изменения концентраций метаболитов биосинтеза валина с помощью кинетической модели в штамме *C. glutamicum* ATCC 13032  $\Delta$ panBC  $\Delta$ ilvA pJC1ilvBNCD показал, что лимитирующими скорость участками в этой цепи являются реакции, катализируемые ферментами АНАС и ВСАТ, и транспорт валина из клеток при помощи BrnFE (Magnus et al., 2009).

### Создание валин-продуцирующих штаммов на основе *C. glutamicum*

Полученная к настоящему времени информация о биохимических, генетических и регуляторных аспектах биосинтеза валина у *C. glutamicum* позволяет считать, что барьерами на пути увеличения продукции валина у этого микроорганизма являются:

- негативная регуляция активности АНАС валином, лейцином, изолейцином и 2-оксоизовалератом (ретроингибирование);
- низкая субстратная специфичность АНАС к пирувату;
- негативная регуляция экспрессии оперона *ilvBNC* при помощи ВСАА;
- расходование пирувата на синтез изолейцина, лейцина и пантотената, а также расходование 2-оксоизовалерата на синтез последних двух соединений;
- расходование пирувата и его предшественника ФЕП – ключевых метаболитов гликолитических процессов – в энергетическом метаболизме клетки и синтезе карбоновых кислот, а также в образовании аланина;
- необходимость НАДФН для второй и четвертой реакций биосинтеза валина;
- низкая эффективность системы экспорта ВСАА BrnFE по отношению к валину.

Далее мы рассмотрим подходы к преодолению этих преград, использованные при создании валин-продуцирующих штаммов на основе *C. glutamicum* (информация о штаммах представлена в таблице).

### Усиление реакций биосинтеза валина

**Увеличение активности АНАС.** Существует несколько подходов к увеличению активности АНАС при создании валин-продуцирующих штаммов. Ключевым является модификация гена *ilvN*, устраняющая ретроингибирование фермента. Известен целый ряд мутаций в последовательности гена *ilvN*, ведущих к ослаблению действия ВСАА на активность АНАС. К таким мутациям относятся замены трех аминокислот, Gly20Asp, Ile21Asp и Ile22Phe, в регуляторной субъединице IlvN (Elišáková et al., 2005). Наилучший эффект в этом ряду показала замена Ile22Phe, которая позже была использована и в других исследованиях (Hou et al., 2012a, b). Аналогичное действие продемонстрировали для мутаций, ведущих к заменам Ala42Val, Ala89Val и Lys136Glu в малой субъединице. Двойная мутация Ala42Val-Ala89Val привела к почти полной устойчивости фермента к ингибированию всеми тремя ВСАА (Guo et al., 2014).

**Усиление субстратной специфичности АНАС к пирувату.** Этот подход связан с возможностью модификаций каталитической субъединицы IlvB АНАС, увеличивающих сродство фермента к пирувату. Достоверных данных о подходящих мутациях мало. Найдена одна

мутация в гене *ilvB*, ведущая к замене аланина на валин в позиции 138 большой субъединицы АНАС, позволившая в 2.5 раза увеличить продукцию валина (Liu et al., 2019). Предполагается, что такая замена приводит к изменению субстратной специфичности АНАС по отношению к пирувату. Молекулярный механизм действия мутации остается неясным.

Известны и другие мутации в гене *ilvB* каталитической субъединицы АНАС, приводящие к увеличению активности фермента в направлении продукции валина (Chen et al., 2015; Guo et al., 2015), которые пока не нашли практического применения.

Для введения в клетки *C. glutamicum* модифицированного фермента АНАС их трансформируют плазмидой, несущей мутантный ген (Hasegawa et al., 2012; Hou et al., 2012b; Buchholz et al., 2013), либо вносят соответствующие изменения в хромосомную ДНК (Bartek et al., 2010; Hasegawa et al., 2013). Подобные манипуляции приводят к увеличению продукции валина в 2–3 раза. Использование автономных экспрессионных плазмид позволяет также увеличивать активность АНАС за счет введения в клетки дополнительных копий генов *ilvBN* или всего оперона *ilvBNC* целиком. Последнее ведет к увеличению активности не только АНАС, но и АНАИР.

**Преодоление негативного влияния ВСАА на экспрессию оперона *ilvBNC*.** Наиболее рациональным подходом к решению данной проблемы является сверхэкспрессия оперона *ilvBNC* с использованием экспрессионных плазмид. В настоящее время сверхэкспрессию генов *ilvBNCDE*, в различных сочетаниях, осуществляют с помощью конструкций с сильными конститутивными промоторами. Таковыми являются, например, промоторы генов супероксиддисмутазы (*Psod*) и фактора элонгации Tu (*Ptuf*) и синтетическая конструкция на основе промоторов оперонов *trp* и *lac* (*Ptac*) (Tarutina et al., 2016; Wei et al., 2018; Li et al., 2020b; Wang et al., 2020; Zhang et al., 2021). Описаны и другие эффективные промоторы (Tarutina et al., 2016; Wei et al., 2018; Li et al., 2020b). Модификации такого типа приводят к увеличению продукции валина примерно на 60 % (Wei et al., 2018).

**Оптимизация активности DHAD и ВСАТ,** катализирующих последние этапы биосинтеза валина, обеспечивается сверхэкспрессией кодирующих эти ферменты генов (*ilvD* и *ilvE* соответственно), которая обычно достигается путем амплификации генов на плаزمиде (см. таблицу). Для продукции валина особенно важно увеличить активность ВСАТ, поскольку этот фермент катализирует скорость-лимитирующий этап биосинтеза (Magnus et al., 2009).

### Минимизация образования побочных продуктов

Ферменты пути биосинтеза валина участвуют в образовании других метаболитов, таких как изолейцин, лейцин и пантотенат (см. рис. 1). Следовательно, активация этих ферментов и усиление экспрессии кодирующих их генов увеличивают выход всех перечисленных соединений, что ведет к загрязнению целевого продукта, а также к уменьшению доступности кофакторов, интермедиатов и самих ферментов для продукции валина и, как результат,

Валин-продуцирующие штаммы, сконструированные на основе *C. glutamicum*

Штамм	Валин, г/л*	Конверсия, моль/моль**	Литературный источник
<b><i>C. glutamicum</i> ATCC 13032</b>			
$\Delta ilvA \Delta panBC$ (pJC1- <i>ilvBNCD</i> ) 92 NA	10.7	–	Radmacher et al., 2002
$\Delta ilvA \Delta panBC ilvNM13$ (pECKA- <i>ilvBNC</i> )	15.2	–	Elišáková et al., 2005
$\Delta panB ilvNM13$ (P- <i>ilvAM1CG</i> P- <i>ilvDM7</i> P- <i>ilvEM6</i> )	15.9	–	Holátko et al., 2009
$\Delta aceE$ (pJC4- <i>ilvBNCE</i> )	24.6	0.60	Blombach et al., 2007
$\Delta aceE \Delta pqr$ (pJC4- <i>ilvBNCE</i> )	26.4	0.52	Blombach et al., 2008
$\Delta aceE \Delta pqr \Delta pgi$ (pJC4- <i>ilvBNCE</i> )	48.2	0.75	»
$\Delta aceE \Delta pqr \Delta pgi \Delta prc$ (pJC4- <i>ilvBNCE</i> )	28.1	0.86	»
$\Delta aceE \Delta pqr$ (pJC4- <i>ilvBNCE</i> )	24.6	0.23	Blombach et al., 2009
$\Delta aceE \Delta pqr \Delta sugR$ (pJC4- <i>ilvBNCE</i> )	35.2	0.20	»
$\Delta aceE \Delta pqr$ (pJC4- <i>ilvBNCE</i> ) (pBB1- <i>pntAB</i> )	14.6	0.92	Bartek et al., 2011
(P- <i>ilvAM1CG</i> ) $\Delta avtA$ pDXW-8- <i>ilvEBN(r)</i> C	31.2	0.17	Hou et al., 2012a
$\Delta ilvA \Delta panB ilvNM13$ (pECKA- <i>ilvBNC</i> )	12.5	–	Denina et al., 2010
$\Delta ilvA \Delta panB \Delta rel ilvNM13$ (pECKA- <i>ilvBNC</i> )	11.5	–	»
$\Delta ilvA \Delta panBC \Delta avtA$ (pJC4- <i>ilvBNCE</i> )	8.8	–	Marienhagen, Eggeling, 2008
$aceEA16 \Delta pqr \Delta prc$ (pJC4- <i>ilvBNCE</i> )	86.5	0.36	Buchholz et al., 2013
$\Delta prc \Delta prc icd^{Ala94Asp}$ (pJC4- <i>ilvBNCE</i> )	8.0	0.20	Schwentner et al., 2018
$\Delta prc \Delta prc icd^{Gly407Ser}$ (pJC4- <i>ilvBNCE</i> )	8.9	0.22	»
$\Delta ponA \Delta ilvA P ilvB^{G183A}$	15.6	–	Рябченко и др., 2021
<b><i>C. glutamicum</i> ATCC 13869</b>			
$\Delta aceE \Delta alaT \Delta ilvA$ (pJYW4- <i>ilvBNC1-lrp1-brnFE</i> )	51.2	0.47	Chen et al., 2015
$\Delta ponA \Delta ilvA P ilvB^{G183A}$	25.1	–	Рябченко и др., 2021
<b><i>C. glutamicum</i> R</b>			
$\Delta ldhA$ (pCRB-BN <sup>GE</sup> C <sup>TM</sup> ) (pCRB-DLD)	172.2 <sup>***</sup>	0.63	Hasegawa et al., 2012
$\Delta ldhA \Delta prc \Delta pta \Delta ackA \Delta ctgA \Delta avtA ilvN^{GE}C^{TM} + gapA+pyk+pfkA+pgi+tpi$ (pCRB-BN <sup>GE</sup> C <sup>TM</sup> ) (pCRB-DLD)	149.9 <sup>***</sup>	0.88	Hasegawa et al., 2013
<b><i>B. flavum</i> ATCC14067 (<i>C. glutamicum</i> ssp. <i>flavum</i>)</b>			
pDXW-8- <i>ilvEBN(r)</i>	38.1	0.24	Hou et al., 2012b
$\Delta alr \Delta aceE \Delta ilvA \Delta leuA$ (pJYW4)	9.5	–	Liu et al., 2019
$\Delta alr \Delta aceE \Delta ilvA \Delta leuA$ (pJYW4- <i>ilvB</i> <sup>138Val404Ala</sup> - <i>ilvN</i> )	14.5	–	»
$\Delta alr \Delta aceE \Delta ilvA \Delta leuA$ (pJYW4- <i>ilvB</i> <sup>138Val404Ala</sup> - <i>ilvNCE</i> )	25.9	0.49	»
<b><i>B. flavum</i> JV16</b>			
<i>avtA::Cm</i> (pDXW8- <i>ilvEBN(r)</i> C)	34.4	0.22	Hou et al., 2012a
<b><i>C. glutamicum</i></b>			
$\Delta prc \Delta aceE \Delta alaT \Delta pqr$	3.2	–	Han et al., 2020

\* Концентрация валина в культуральной жидкости.

\*\* Выход целевого продукта (валина) из субстрата (глюкозы).

\*\*\* Данные получены с использованием концентрированной суспензии клеток.

к снижению уровня конверсии. Минимизация образования побочных продуктов при создании штаммов-продуцентов требует подавления соответствующих метаболических путей при сохранении способности штаммов расти на бедных средах.

**Минимизация образования изолейцина.** Выше отмечалось, что синтез изолейцина (2-амино-3-метилпен-

тановой кислоты) катализируется теми же ферментами, какие участвуют в биосинтезе валина, и начинается с конденсации пирувата и 2-оксобутирата (см. рис. 1). Очевидный способ минимизации образования изолейцина – снижение концентрации в клетках 2-оксобутирата, взаимодействием которого с пируватом определяется направление дальнейших реакций. 2-Оксобутират образуется

из треонина под действием треониндегидратазы TD, которая кодируется геном *ilvA* (Cordes et al., 1992). Треониндегидратаза подвержена негативной аллостерической регуляции изолейцином и положительной – валином (Möckel et al., 1992).

Наиболее распространенной модификацией данного гена при создании валин-продуцирующих штаммов является его инактивация путем делетирования ( $\Delta ilvA$ ). С использованием такой модификации получено большинство штаммов (см. таблицу). Результатом ее становится появление у штаммов способности продуцировать валин либо значительное увеличение уже имеющейся продукции. При этом возникает ауксотрофия по изолейцину, требующая добавления изолейцина в культивационную среду, что усложняет производственный процесс и может повысить себестоимость продукции. В ряде исследований для создания валин-продуцирующих штаммов вместо полной инактивации гена *ilvA* произвели направленную модификацию его промотора, приводящую к снижению экспрессии гена, появлению брадитрофии по изолейцину и, как следствие, повышению продукции валина (Holátko et al., 2009; Hou et al., 2012a).

Еще одна мишень для модификаций, направленных на снижение биосинтеза изолейцина, – это ANAS. Вариант модификации каталитической субъединицы фермента, увеличивающий его специфичность по отношению к пирувату и перенаправляющий ресурсы клетки на продукцию валина (Liu et al., 2019), описан выше.

**Минимизация образования лейцина и пантотената.** 2-Оксоизовалерат является предшественником не только валина, но и лейцина и пантотената (см. рис. 1). Синтез лейцина (2-амино-4-метилпентаановой кислоты) из 2-оксоизовалерата контролируется генами *leuA*, *leuB* и *leuCD*, локализованными в разных участках хромосомы. Известно, что *leuB* и *leuCD* подвержены контролю транскрипционного репрессора LtbR, а в регуляции *leuA*, по-видимому, участвует механизм аттенуации транскрипции (Wang et al., 2019a). Модификацию, направленную на сохранение 2-оксоизовалерата для биосинтеза валина за счет уменьшения биосинтеза лейцина, осуществили J. Holátko с коллегами (2009), снизив экспрессию гена *leuA* путем замены нативного промотора более слабым. Результатом стало увеличение продукции валина на 50–70 %, что сравнимо с эффектом, наблюдаемым при ослаблении экспрессии гена *ilvA*.

Синтез пантотената (амида  $\beta$ -аланина и пантоевой кислоты) из 2-оксоизовалерата контролируется генами *panB* и *panC*, которые образуют один оперон (Sahm, Eggeling, 1999), и геном *panD*, расположенным в геноме отдельно (Dusch et al., 1999). Отмечено, что поток углерода, идущий на биосинтез валина, в 10 раз выше потока, идущего на биосинтез пантотената, даже в штамме с усиленной экспрессией *panBC* (Chassagnole et al., 2002). Тем не менее инактивация гена *panB* или всего оперона *panBC* хоть и ведет к ауксотрофии штаммов по пантотенату, но позволяет добиться появления продукции валина у штаммов дикого типа и увеличить продукцию валина у валин-продуцирующих штаммов более чем на 30 и даже 50 % (Radmacher et al., 2002; Holátko et al., 2009).

## Увеличение доступности предшественников и кофакторов

**Увеличение доступности пирувата.** Пируват, центральное звено углеродного и энергетического обмена во всех организмах, – предшественник не только ВСАА и пантотената, но и множества других соединений, включая компоненты цикла трикарбоновых кислот (ЦТК), а также лактат и аланин (см. рис. 1). Эффективная продукция валина требует поддержания в клетках пула пирувата, а, следовательно, усиления реакций образования пирувата и/или уменьшения его нецелевого расхода. Пируват, при образовании которого синтезируется 2 моля восстановленного никотинамиддинуклеотида (НАДН), является продуктом гликолиза (Wieschalka et al., 2012). Однако при разработке продуцентов валина активуют гликолитических ферментов обычно не увеличивают, за исключением микроаэробного процесса (см. ниже). Основной подход – снижение оттока пирувата и его предшественника ФЕП в другие пути.

Одно из главных направлений оттока пирувата – ЦТК. Данный процесс становится менее активным на поздних стадиях роста, что можно было бы использовать при создании благоприятных условий для продукции валина. Действительно, снижение скорости роста *C. glutamicum* сопровождается повышением уровня пирувата в клетках и увеличением количества валина (Ruklisha et al., 2007). В валин-продуцирующих штаммах, ауксотрофных по изолейцину и пантотенату, рост культур можно контролировать, изменяя количество подпитки этими веществами; ограничение роста также ведет к увеличению продуктивности (Bartek et al., 2008).

Вовлечение пирувата и ФЕП в ЦТК происходит в результате превращения как обоих соединений в оксалоацетат (ОА), так и пирувата в ацетил-кофермент А (ацетил-КоА) непосредственно или через ацетат и ацетил-фосфат (см. рис. 1). Как правило, повышения продукции валина, а также самого пирувата пытаются добиться через снижение активности пируватдегидрогеназного комплекса (PDHC), который катализирует окислительное декарбоксилирование пирувата до ацетил-КоА. У *C. glutamicum* данный комплекс состоит из трех субъединиц, E1, E2 и E3, кодируемых генами *aceE*, *aceF* и *lpd* соответственно (Eikmanns, Blombach, 2014). Инактивация гена *aceE* делетированием ( $\Delta aceE$ ) – один из наиболее частых шагов при создании продуцента валина (см. таблицу). Полученные штаммы требуют добавления ацетата в минимальную среду, но уровень продукции валина возрастает многократно. Метаболомный анализ показал, что инактивация *aceE* у *C. glutamicum* дикого типа ведет к 13-кратному увеличению пула пирувата в клетках (Blombach et al., 2007).

Характерная особенность штаммов *C. glutamicum*, лишенных PDHC, – продукция валина в отсутствие роста клеток. Увеличения скорости утилизации глюкозы добивались добавлением мальтозы вместо глюкозы, использованием этанола вместо ацетата или инактивацией регулятора транскрипции SugR (Blombach et al., 2009; Krause et al., 2010b). SugR у *C. glutamicum* отвечает за ацетат-опосредованную репрессию генов *ptsG*, *ptsI* и *ptsH*, кодирующих ферменты фосфотрансферной системы (PTS), которая

обеспечивает сопряженные процессы транспорта сахаров в клетку и их фосфорилирования (Engels, Wendisch, 2007). Однако из-за дефицита PDHC все штаммы по-прежнему нуждались в ацетате или этаноле, который затем также превращается в ацетат в качестве дополнительного источника углерода.

Для преодоления этой потребности провели замену нативного промотора гена *aceE* на мутантные варианты из ранее созданной библиотеки промоторов на основе промотора гена *dapA* (Vasicová et al., 1999), что позволило получить серию штаммов *C. glutamicum* с постепенным снижением активности PDHC и скорости роста на среде, содержащей глюкозу в качестве единственного источника углерода. Трансформированные плазмидой pJC4-*ilvBNCE*, данные штаммы продуцировали валин и не требовали ацетата в качестве дополнительного источника углерода (Buchholz et al., 2013). Использование для тех же целей промотора гена альдегиддегидрогеназы из *C. glutamicum* CP (P<sub>CP</sub><sub>2836</sub>), идентифицированного как промотор, регулируемый ростом, и снижающего уровень транскрипции *aceE* втрое по сравнению с нативным промотором, также хорошо сказалось и на росте клеток, и на продукции валина (Ma et al., 2018b).

Уменьшить расход пирувата в ЦТК можно и за счет снижения активности самого цикла. Так, оказалось, что эффективной продукции пирувата способствует подавление гена транскрипционного фактора RamA, ответственного за активацию ЦТК (Kataoka et al., 2019).

Превращение пирувата в ацетат катализируется пируват:хиноноксидоредуктазой (продукт гена *pqo*), инактивация которой ( $\Delta pqo$ ) ведет к повышению продукции валина (см. таблицу), но и к ухудшению ростовых характеристик штаммов. Объединение такой модификации с инактивацией ФЕП-карбоксилазы (продукт гена *ppc*), катализирующей образование ОА из ФЕП, привело к незначительному повышению продукции валина, однако показатель конверсии глюкозы вырос на 14 % (Buchholz et al., 2013). Отмечается, что валин-продуцирующий штамм с инактивированными генами *aceE* и *pqo* лучше растет и продуцирует больше валина на среде, обогащенной мальтозой (Krause et al., 2010b).

Еще один путь оттока пирувата – образование из него ОА под действием пируваткарбоксилазы (продукт гена *pus*). Инактивация *pus* при создании продуцента валина приводит к увеличению конверсии глюкозы до 0.86 моль валина на 1 моль глюкозы (Blombach et al., 2008). При разработке лейцин-продуцирующего штамма выяснилось, что для накопления пирувата инактивация пируваткарбоксилазы выгодней, чем инактивация ФЕП-карбоксилазы (Wang et al., 2020).

Два других пути расходования пирувата в клетках *C. glutamicum* – процессы биосинтеза лактата и аланина (см. рис. 1). Образование лактата, катализируемое лактатдегидрогеназой (продукт гена *ldhA*), приобретает значение с точки зрения продукции валина в условиях недостатка кислорода (Hasegawa et al., 2012) и будет рассмотрено далее.

Минимизация синтеза аланина требуется в любых условиях, поскольку этот процесс ведет не только к нецелевому расходу пирувата, но также к потере НАДФН в реакции

переноса аминогруппы и к появлению нежелательных примесей в конечном продукте.

Образование аланина у коринебактерий катализируют трансаминазы AlaT и AvtA, использующие в качестве доноров аминокрупп глутамат и валин соответственно (Marienhagen et al., 2005; Marienhagen, Eggeling, 2008). Выше отмечалось, что AvtA является одной из трансаминаз, участвующих в биосинтезе валина, но ее роль, по сравнению с ВСАТ, незначительна.

Вопрос об участии этих трансаминаз в биосинтезе аланина у *C. glutamicum* остается открытым в силу противоречивости существующих данных. С одной стороны, показано, что инактивация *alaT* и *avtA* в валин-продуцирующем штамме снижает образование аланина примерно на 80 и 20 % соответственно (Marienhagen, Eggeling, 2008). Значительное снижение синтеза аланина (до уровня менее 0.2 г/л) наблюдается в результате инактивации обоих генов (Hou et al., 2012a). Эти данные позволяют сделать вывод, что аминотрансфераза AlaT является основной, но в синтезе аланина участвуют оба фермента. С другой стороны, в продуценте пролина инактивация *alaT* не влияет на уровень аланина, тогда как инактивация *avtA* снижает этот уровень на 48 % (Zhang et al., 2020). Более того, анализ транскриптома промышленного продуцента валина линии VWB-1 показал, что низкий уровень синтеза аланина в ней не связан с геном *alaT*, уровень транскрипции которого в этом штамме в 5.1 раза выше, чем таковой в штамме дикого типа. Предполагается, что снижение образования аланина связано с более низкой экспрессией гена *alr*, кодирующего аланинрацемазу, конвертирующую L-аланин в D-аланин (Zhang H. et al., 2018). Таким образом, нельзя дать однозначный ответ и на вопрос о том, инактивация которой из этих двух трансаминаз более выгодна с точки зрения продукции валина.

**Увеличение доступности НАДФН.** У *C. glutamicum* основным поставщиком НАДФН является ПФП, в котором восстановление НАДФ<sup>+</sup> до НАДФН обеспечивается глюкозо-6-фосфатдегидрогеназой (гетеромультимерный комплекс, одна из субъединиц которого кодируется геном *zwf*) и 6-фосфоглюконатдегидрогеназой (продукт гена *gnd*). Активность обоих ферментов подвержена негативной регуляции со стороны АТФ, НАДФН и других метаболитов (Moritz et al., 2000). Малик-фермент (НАДФ-зависимая декарбоксилирующая малатдегидрогеназа) и изоцитратдегидрогеназа играют минорную роль в синтезе НАДФН из НАДФ<sup>+</sup> (Bartek et al., 2010; Siedler et al., 2013). Источником НАДФ<sup>+</sup> и, следовательно, НАДФН у *C. glutamicum* также может быть НАД<sup>+</sup>, который фосфорилируется НАД-киназой (продукт гена *ppnK*) с образованием НАДФ<sup>+</sup>. Этот фермент охарактеризован как полифосфат-АТФ-зависимая НАД-киназа, использующая АТФ для фосфорилирования НАД<sup>+</sup> (Shi et al., 2013).

Теоретический анализ показал, что от реакций, используемых для регенерации НАДФН, существенно зависит уровень конверсии субстрата в валин. Максимальный уровень конверсии, равный 1 молью валина на 1 моль глюкозы, получается без расходования углерода на рост и синтез НАДФН. Если НАДФН обеспечивается активностью изоцитратдегидрогеназы, то показатель конверсии равен 0.5 моля валина на 1 моль глюкозы. Направление всего

потока углерода в ПФП, генерирующий НАДФН, приводит к значительно более высокой конверсии – 0.86. В этом анализе главной мишенью для перенаправления потока углерода из ЦТК в биосинтез валина оказался PDHC. Сценарий, при котором углерод вообще не расходуется на синтез НАДФН, может быть реализован за счет комбинированной активности пируваткарбоксилазы (или ФЕП-карбоксилазы), малатдегидрогеназы и малик-фермента, теоретически способного переносить водород от НАДН к НАДФ<sup>+</sup> (Bartek et al., 2010). Такой путь, обозначенный как трансгидрогеназоподобный шунт, участвует в образовании НАДФН для анаэробной продукции изобутанола у *C. glutamicum* (Blombach, Eikmanns, 2011). Таким образом, усиление ПФП и активности НАД-киназы – наиболее очевидные способы увеличения пула НАДФН в клетке.

С точки зрения эффективности процесса биосинтеза валина усиление ПФП выгодно сочетать с некоторым ослаблением гликолиза. В самом деле, инактивация гена глюкозо-6-фосфатизомеразы *pgi*, направляющая поток углерода от гликолиза в ПФП, привела к более эффективной продукции валина в штамме *C. glutamicum*  $\Delta aceE \Delta pgi \Delta pglvBNCE$  – образованию 48.0 г/л с конверсией 0.75 моль валина на 1 моль глюкозы (Blombach et al., 2008). Дальнейший анализ этого штамма показал, что инактивация *pgi* приводит к повышению внутриклеточных концентраций НАДФН и снижению образования побочных продуктов (Bartek et al., 2010). Мониторинг содержания НАДНФ в клетке с помощью НАДФН-зависимой флуоресценции также показал, что штамм *C. glutamicum*, несущий  $\Delta pgi$ , действительно накапливает НАДФН (Goldbeck et al., 2018).

Ухудшение роста, наблюдающееся у  $\Delta pgi$ -штаммов на среде с глюкозой, связывают со снижением активности PTS и предлагают преодолевать сверхэкспрессией гена *ptsG*, кодирующего глюкозо-специфичный компонент этой системы (Lindner et al., 2013). Для *pgi*-мутантов также эффективно усиление альтернативной системы транспорта глюкозы с помощью инозитолпермеаз IolT1, IolT2 и глюкокиназы PrgK, которую использовали для получения продуцента лизина (Xu J.Z. et al., 2019).

Еще один подход к увеличению пула НАДФН связан с возможностью изменения специфичности гликолитических ферментов с НАД<sup>+</sup> на НАДФ<sup>+</sup>. Он был реализован для улучшения продукции лизина – точечные мутации в гене глицеральдегид-3-фосфатдегидрогеназы *gapA*, изменившие специфичность фермента, привели к повышению продукции лизина на 35–60 % (Vommarreddy et al., 2014; Xu et al., 2014).

Выше отмечено, что ферменты, синтезирующие НАДФН, подвержены негативной регуляции различными метаболитами. Поэтому одним из подходов к активации ПФП является внесение в соответствующие гены мутаций, повышающих активность ферментов. Такой подход был реализован для генов *zwf* и *gnd* в работах по созданию продуцентов метионина, пролина и рибофлавина и действительно привел к увеличению в клетках пула НАДФН и повышению уровня продукции (Wang et al., 2011; Li et al., 2016; Zhang et al., 2020).

Что касается НАД-киназы, то опубликованные к настоящему времени исследования по усилению ее активности

нацелены на продукцию изолейцина и свидетельствуют о том, что модификации, увеличивающие активность фермента (точечные мутации в гене *ppnK*, сверхэкспрессия гена *ppnK*), ведут к повышению внутриклеточной концентрации НАДФ<sup>+</sup> и НАДФН и способствуют усилению биосинтеза целевого продукта (Yin et al., 2014; Zhang et al., 2020).

Другой привлекательной возможностью увеличения доступности НАДФН для биосинтеза валина является гетерологичная экспрессия генов трансгидрогеназ, таких как *pntAB* из *E. coli*, катализирующих восстановление НАДФ<sup>+</sup> с участием НАДН. Ранее эта возможность использовалась для улучшения продукции лизина с помощью *C. glutamicum* (Kabus et al., 2007). Значительное увеличение внутриклеточной концентрации НАДФН наблюдалось при сочетании экспрессии *pntAB* со сверхэкспрессией гена *ppnK* (Zhan et al., 2019). Введение PntAB из *E. coli* в валин-продуцирующий штамм *C. glutamicum*  $\Delta aceE \Delta pgi$  (pJC4*ilvBNCE*) привело к существенному снижению потока углерода в ПФП и, соответственно, к увеличению конверсии до 0.92. Это самый высокий показатель конверсии (Bartek et al., 2011), который всего на 8 % ниже теоретического максимума, равного 1 моль валина на 1 моль глюкозы (Bartek et al., 2010).

### Инженерия микроаэробного процесса продукции валина

При недостатке кислорода культуры *C. glutamicum* проявляют очень слабую способность к росту, однако метаболизируют сахара до органических кислот (Michel et al., 2015; Lange et al., 2018). При подавлении синтеза побочных продуктов штаммы-продуценты, адаптированные к таким условиям, показывают более высокую продуктивность, чем штаммы, требующие аэрации (Okino et al., 2008; Jojima et al., 2010, 2015; Yamamoto et al., 2013). Биосинтез валина в обычных условиях является аэробным процессом, так как осуществляется растущими культурами, активно генерирующими НАДФН. Для эффективной продукции валина при недостатке кислорода штаммам требуется комплексная модификация, затрагивающая ферменты как биосинтеза валина, так и гликолиза; такую модификацию осуществили S. Hasegawa с коллегами (2012, 2013).

В качестве базы для создания штаммов, продуцирующих валин в микроаэробных условиях, использовали штамм *C. glutamicum* R с инактивированной лактатдегидрогеназой ( $\Delta ldhA$ ) и сверхэкспрессией генов *ilvBNCE*, кодирующих ферменты биосинтеза валина. Данный штамм не способен к продукции валина при недостатке кислорода, поскольку имеет дисбаланс кофакторов: при образовании 1 моля валина расходуется 2 моля НАДФН, а синтезируется 2 моля НАДН.

Появления продукции валина удалось достичь благодаря использованию двух подходов. Первый – изменение специфичности АНАIR с НАДФН на НАДН с помощью сайт-направленного мутагенеза гена *ilvC* (конструирование гена *ilvC<sup>TM</sup>*), а второй – замена НАДФН-зависимой трансминазы BCAT на НАД-зависимую лейциндегидрогеназу (LeuDh) из *Lysinibacillus sphaericus* (Hasegawa et al., 2012). Дополнительное введение гена *ilvN*, кодирующего мутантную регуляторную субъединицу АНАS (*ilvN<sup>GE</sup>*),

устойчивую к ингибированию ВСАА, позволило создать штамм *C. glutamicum* (pCRB-BN<sup>GE</sup>C<sup>TM</sup>)(pDLD)/ΔLDH), который в микроаэробных условиях при периодической ферментации с подпиткой продуцировал 172.2 г/л валина в течение 24 ч, что превышало исходный показатель более чем в 20 раз. Уровень конверсии составил 0.63 моль валина на моль глюкозы (Hasegawa et al., 2012).

Однако, помимо валина, клетки данного штамма накапливали значительные количества аланина, ацетата и сукцината в качестве побочных продуктов. Чтобы исключить их образование и повысить выход валина, в штамм были введены дополнительные модификации (Hasegawa et al., 2013). Образование сукцината через ФЕП и ОА подавили инактивацией гена *ppc*, но это привело к снижению синтеза валина и поглощения глюкозы, так как внутриклеточное соотношение НАДН/НАД<sup>+</sup> заметно повысилось. Чтобы вернуть соотношение на благоприятный для продукции валина уровень, инактивировали три гена, участвующих в синтезе ацетата (*pta*, *ackA*, *ctfA*), и усилили экспрессию пяти генов (*gapA*, *pyk*, *pfkA*, *pgi*, *tpi*), кодирующих ферменты гликолиза. В результате продукция валина выросла в 9 раз, поглощение глюкозы – в 7.6 раза. Поскольку биосинтез валина стал НАДН-зависимым процессом, увеличение активности гликолитических ферментов оказалось выгодным с точки зрения накопления как пирувата, так и восстановительных эквивалентов.

Уменьшения образования аланина добились инактивацией гена *avtA*. Кроме того, гены *ilvN<sup>GE</sup>* и *ilvC<sup>TM</sup>*, которые ранее экспрессировались на плазмиде, были помещены в хромосому. Продуктивность нового штамма по валину составила 149.9 г/л за 24 ч культивирования. Конверсия достигла 0.88 моля валина на моль глюкозы, что значительно превысило показатели, полученные на первом этапе (Hasegawa et al., 2013).

Следует отметить, что в обеих работах синтез валина в микроаэробных условиях изучали с использованием не-растущих клеток, предварительно сконцентрированных центрифугированием в 2–3 раза. В этом случае измеренная концентрация валина достигала очень высоких значений, но продуктивность на клетку была сопоставимой с продуктивностью, продемонстрированной в других исследованиях.

Замена специфичности ферментов с НАДФН на НАДН для адаптации процесса производства аминокислот к микроаэробным условиям была осуществлена также при разработке продуцента валина на базе *E. coli* (Savrasova, Stoyanova, 2019) и продуцентов лейцина и L-орнитина на базе *C. glutamicum* (Jiang et al., 2013; Wang et al., 2019b). Во всех случаях это привело к увеличению выхода целевого продукта.

## Инженерия транспорта валина

Микроорганизмы располагают множеством транспортных систем, обеспечивающих поглощение клетками нужных компонентов среды и освобождение от метаболитов, избыток которых может оказаться токсичным (Pérez-García, Wendisch, 2018). Активность таких систем зависит от концентрации транспортируемых веществ, поэтому долгое время считалось, что для эффективного вывода целевых продуктов штаммам-продуцентам достаточно соб-

ственных регуляторных механизмов (Jones et al., 2015). Инженерию транспорта затрудняют сложность его количественной оценки и тот факт, что специфические транспортеры известны не для каждого из биотехнологически значимых веществ. Однако в последние годы растет число исследований, показавших влияние направленного изменения экспорта и импорта целевого продукта на продуктивность штаммов (Eggeling, 2016). Транспортеры валина у коринебактерий обнаружены и охарактеризованы, а значит, являются перспективным объектом модификаций при создании продуцентов.

**Импорт валина.** Поглощение валина, лейцина и изолейцина у коринебактерий происходит в ходе вторичного Na<sup>+</sup>-зависимого симпорта, осуществляемого при помощи единственного известного импортера – BrnQ (Ebbighausen et al., 1989). Наибольшее сродство BrnQ проявляет к изолейцину, для валина и лейцина этот показатель в 1.7 раза ниже (Ebbighausen et al., 1989; Tauch et al., 1998). Данных о регуляции BrnQ и соответствующего гена у коринебактерий крайне мало. Известно, что BrnQ активируется при повышении внутриклеточной концентрации ВСАА (Boles et al., 1993) и что инактивация гена *brnQ* увеличивает экспорт изолейцина из клеток *C. glutamicum* и его продукцию (Xie et al., 2012). Отмечается, что аналогичная модификация благоприятствует росту и продуктивности изолейцин-продуцирующего штамма WM001 на ранних стадиях ферментации (Zhang et al., 2020). Значение импортера для продукции валина подтверждается транскриптомным анализом промышленного продуцента VWB-1, показавшим, что уровень транскрипции гена *brnQ* в этом штамме ниже, чем в штамме дикога типа (Zhang H. et al., 2018).

**Экспорт валина.** За выведение ВСАА из клеток *C. glutamicum* отвечает транспортная система BrnFE (Eggeling, Sahm, 2003). Аминокислоты экспортируются в ходе вторичного H<sup>+</sup>-зависимого процесса, управляемого мембранным потенциалом (Hermann, Kramer, 1996). Транспортная система BrnFE – единственный известный экспортер валина, лейцина и изолейцина у *C. glutamicum*; также она переносит метионин и гомосерин – предшественник метионина, изолейцина и треонина (Kennerknecht et al., 2002; Trotschel et al., 2005; Yin et al., 2013; Qin et al., 2015; Li et al., 2020a). Гены *brnF* и *brnE*, кодирующие, соответственно, большую и малую субъединицы транспортной системы, организованы в один оперон, контролируемый транскрипционным регулятором Lrp (leucine responsive protein) (Kennerknecht et al., 2002; Lange et al., 2012). Гомологи Lrp, впервые обнаруженного и охарактеризованного в *E. coli*, присутствуют в геномах различных прокариот и регулируют гены, задействованные в метаболизме аминокислот (Brinkman et al., 2003). У *C. glutamicum* ген *lrp* расположен дивергентно перед опероном *brnFE*. Связываясь с ВСАА или метионином, Lrp переходит в активное состояние и в свою очередь активирует промотор *brnFE* (Kennerknecht et al., 2002; Lange et al., 2012) (рис. 3). Влияние концентрации аминокислот в клетке на активность Lrp уменьшается в ряду лейцин > метионин > изолейцин > валин (Lange et al., 2012).

Изучение промышленных продуцентов лейцина и валина подтверждает, что высокий уровень продукции аминокислот коррелирует с высоким уровнем экспрессии

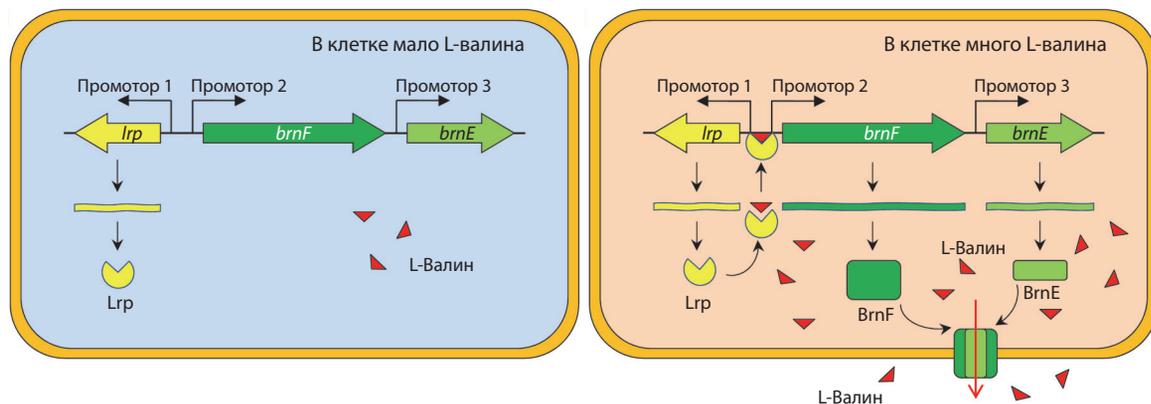


Рис. 3. Организация оперона *brnFE* у *C. glutamicum* и регуляция его экспрессии (из обзора (Wang et al., 2018)).

оперона *brnFE* (Vogt et al., 2014; Zhang H. et al., 2018) или связан с увеличением количества копий генов *lrp* и *brnFE* (Ma et al., 2018a).

Анализ влияния оперона *brnFE* на продукцию валина показал, что его сверхэкспрессия не сказывается на росте клеток *C. glutamicum* и увеличивает продукцию валина примерно в 2–3 раза (Chen et al., 2015). Аналогичный эффект сверхэкспрессия *brnFE* оказывает на продукцию изолейцина, метионина и гомосерина (Qin et al., 2015; Li et al., 2020a; Zhang et al., 2021). Максимальный эффект на продукцию изолейцина был получен при одновременном усилении экспрессии *lrp* и *brnFE* (Yin et al., 2013).

Однако выяснилось, что, в отличие от *brnFE*, сверхэкспрессия гена *lrp* подавляет рост клеток (Chen et al., 2015), хотя также существенно повышает продукцию валина. Негативный эффект был нивелирован использованием ослабленной мутантной формы данного гена *lrp*<sub>1</sub>, обнаруженной у штамма VWB-1. Сверхэкспрессия *lrp*<sub>1</sub> в штамме *C. glutamicum* дикого типа привела к 16-кратному увеличению продукции валина – с 1.9 до 30.2 ммоль/л за 96 ч культивирования. Сочетание сверхэкспрессии *lrp*<sub>1</sub> и *brnFE* усилило эффект. На продукции изолейцина подобные манипуляции сказались незначительно, из чего авторы исследования заключили, что изолейцин – менее подходящий субстрат для BrnFE, чем валин (Chen et al., 2015). Одновременное усиление экспрессии обоих генов, *lrp* и *brnFE*, в сочетании со сверхэкспрессией генов *ilvBNC* и инактивацией *aceE*, *alaT* и *ilvA*, позволило создать штамм, который продуцировал 437 мМ (51 г/л) валина при ферментации с подпиткой (Chen et al., 2015).

Таким образом, модификации систем транспорта ВСАА, направленные на уменьшение притока аминокислот в клетку и увеличение их секреции из клетки, позитивно влияют на продукцию этих аминокислот (Xie et al., 2012).

### Заключение

В последние годы интерес к использованию валина в качестве кормовой добавки существенно вырос. Только в Российской Федерации потребление валина за последние пять лет увеличилось почти в 10 раз и достигло 5000 т в год. Современное промышленное производство валина

базируется на микробиологическом синтезе с использованием возобновляемого растительного сырья и штаммов-продуцентов с измененной генетической программой. Эффективность производства аминокислот в большей степени зависит от продуктивности штаммов-продуцентов, которые являются ключевым элементом всей технологической цепочки. Несмотря на значительный прогресс в создании штаммов-продуцентов (см. таблицу), создание новых штаммов с уникальными характеристиками по-прежнему актуально.

Обращает на себя внимание высокий потенциал процессов со сниженной аэрацией, разработанных в последнее время, по сравнению с традиционными аэробными процессами продукции валина. Однако необходимо отметить, что такие процессы являются двухфазными: в первой фазе аэробно нарабатывается биомасса, во второй происходит биосинтез валина в микроаэробных условиях. В настоящее время двухфазные процессы показывают низкую эффективность, требуются дополнительные исследования в этой области.

Сейчас основным подходом к созданию валин-продуцирующих штаммов, пришедшим на смену случайному мутагенезу, является рациональная метаболическая инженерия, направленная на усиление процесса биосинтеза валина и минимизацию образования побочных продуктов. В последние годы этот подход активно обогащают применением методов системной инженерии и синтетической биологии. Комбинированный анализ «омических» данных расширяет наши знания о метаболических и регуляторных процессах *C. glutamicum* и позволяет вырабатывать новые стратегии создания продуцентов валина и других аминокислот. Реализации этих стратегий должны помочь появившиеся недавно системы быстрого редактирования генома, ускоряющие процесс получения новых штаммов.

Дальнейший прогресс в создании штаммов-продуцентов будет связан с переходом от изучения свойств популяции клеток к изучению свойств отдельных клеток (Harst et al., 2017; Hemmerich et al., 2018; Pérez-García et al., 2018), а также с широким применением компьютерного моделирования (Koduru et al., 2018) и использованием новых знаний о регуляции экспрессии генов (Dostálová et al., 2017; Shi et al., 2018; Zhang S. et al., 2018; Xu N. et al., 2019).

Подходы, отработанные при создании и совершенствовании продуцентов валина, могут быть использованы для создания продуцентов других ВСАА и пантотената – веществ, которые также обладают значительным рыночным потенциалом.

## Список литературы / References

- Рябченко Л.Е., Герасимова Т.В., Леонова Т.Е., Калинина Т.И., Шереметьева М.Е., Ануфриев К.Э., Яненко А.С. Пат. RU 2753996 С1. Бактерия *Corynebacterium glutamicum* с повышенной способностью продуцировать L-валин и способ получения L-валина с использованием этой бактерии. Опул. 25.08.2021. Бюл. № 24. [Ryabchenko L.E., Gerasimova T.V., Leonova T.E., Kalinina T.I., Sheremetieva M.E., Anufriev K.E., Yanenko A.S. Patent RU 2753996 C1. Bacterium *Corynebacterium glutamicum* with increased ability to produce L-valine and method for producing L-valine using this bacterium. Date of publication: 25.08.2021. Bull. No. 24. (in Russian)]
- Baritugo K.A., Kim H.T., David Y., Choi J.I., Hong S.H., Jeong K.J., Choi J.H., Joo J.C., Park S.J. Metabolic engineering of *Corynebacterium glutamicum* for fermentative production of chemicals in biorefinery. *Appl. Microbiol. Biotechnol.* 2018;102(9):3915-3937. DOI 10.1007/s00253-018-8896-6.
- Bartek T., Blombach B., Lang S., Eikmanns B.J., Wiechert W., Oldiges M., Noh K., Noack S. Comparative C-13 metabolic flux analysis of pyruvate dehydrogenase complex-deficient, L-valine-producing *Corynebacterium glutamicum*. *Appl. Environ. Microbiol.* 2011; 77(18):6644-6652. DOI 10.1128/aem.00575-11.
- Bartek T., Blombach B., Zonnchen E., Makus P., Lang S., Eikmanns B.J., Oldiges M. Importance of NADPH supply for improved L-valine formation in *Corynebacterium glutamicum*. *Biotechnol. Prog.* 2010;26(2):361-371. DOI 10.1002/btpr.345.
- Bartek T., Makus P., Klein B., Lang S., Oldiges M. Influence of L-isoleucine and pantothenate auxotrophy for L-valine formation in *Corynebacterium glutamicum* revisited by metabolome analyses. *Bioprocess Biosyst. Eng.* 2008;31(3):217-225. DOI 10.1007/s00449-008-0202-z.
- Blombach B., Arndt A., Auchter M., Eikmanns B.J. L-valine production during growth of pyruvate dehydrogenase complex deficient *Corynebacterium glutamicum* in the presence of ethanol or by inactivation of the transcriptional regulator SugR. *Appl. Environ. Microbiol.* 2009;75(4):1197-1200. DOI 10.1128/aem.02351-08.
- Blombach B., Eikmanns B.J. Current knowledge on isobutanol production with *Escherichia coli*, *Bacillus subtilis* and *Corynebacterium glutamicum*. *Bioeng. Bugs.* 2011;2(6):346-350. DOI 10.4161/bbug.2.6.17845.
- Blombach B., Schreiner M.E., Bartek T., Oldiges M., Eikmanns B.J. *Corynebacterium glutamicum* tailored for high-yield L-valine production. *Appl. Microbiol. Biotechnol.* 2008;79(3):471-479. DOI 10.1007/s00253-008-1444-z.
- Blombach B., Schreiner M.E., Holátko J., Bartek T., Oldiges M., Eikmanns B.J. (L)-valine production with pyruvate dehydrogenase complex-deficient *Corynebacterium glutamicum*. *Appl. Environ. Microbiol.* 2007;73(7):2079-2084. DOI 10.1128/aem.02826-06.
- Boles E., Ebbighausen H., Eikmanns B., Krämer R. Unusual regulation of the uptake system for branched-chain amino acids in *Corynebacterium glutamicum*. *Arch. Microbiol.* 1993;159:147-152. DOI 10.1007/BF00250275.
- Bommareddy R.R., Chen Z., Rappert S., Zeng A.P. A *de novo* NADPH generation pathway for improving lysine production of *Corynebacterium glutamicum* by rational design of the coenzyme specificity of glyceraldehyde 3-phosphate dehydrogenase. *Metab. Eng.* 2014;25: 30-37. DOI 10.1016/j.ymben.2014.06.005.
- Brinkman A.B., Ettema T.J., de Vos W.M., van der Oost J. The Lrp family of transcriptional regulators. *Mol. Microbiol.* 2003;48(2): 287-294. DOI 10.1046/j.1365-2958.2003.03442.x.
- Buchholz J., Schwentner A., Brunnenkan B., Gabris C., Grimm S., Gerstmeir R., Takors R., Eikmanns B.J., Blombach B. Platform engineering of *Corynebacterium glutamicum* with reduced pyruvate dehydrogenase complex activity for improved production of L-lysine, L-valine, and 2-ketoisovalerate. *Appl. Environ. Microbiol.* 2013;79(18):5566-5575. DOI 10.1128/AEM.01741-13.
- Burkovski A. I do it my way: regulation of ammonium uptake and ammonium assimilation in *Corynebacterium glutamicum*. *Arch. Microbiol.* 2003;179(2):83-88. DOI 10.1007/s00203-002-0505-4.
- Chassagnole C., Létisse F., Diano A., Lindley N.D. Carbon flux analysis in a pantothenate overproducing *Corynebacterium glutamicum* strain. *Mol. Biol. Rep.* 2002;29(1-2):129-134. DOI 10.1023/a:1020353124066.
- Che L., Xu M., Gao K., Wang L., Yang X., Wen X., Xiao H., Li M., Jiang Z. Mammary tissue proteomics in a pig model indicates that dietary valine supplementation increases milk fat content via increased *de novo* synthesis of fatty acid. *Food Sci. Nutr.* 2021;9(11): 6213-6223. DOI 10.1002/fsn3.2574.
- Chen C., Li Y., Hu J., Dong X., Wang X. Metabolic engineering of *Corynebacterium glutamicum* ATCC13869 for L-valine production. *Metab. Eng.* 2015;29:66-75. DOI 10.1016/j.ymben.2015.03.004.
- Chen X.H., Liu S.R., Peng B., Li D., Cheng Z.X., Zhu J.X., Zhang S., Peng Y.M., Li H., Zhang T.T., Peng X.X. Exogenous L-valine promotes phagocytosis to kill multidrug-resistant bacterial pathogens. *Front. Immunol.* 2017;8:207. DOI 10.3389/fimmu.2017.00207.
- Cordes C., Möckel B., Eggeling L., Sahn H. Cloning, organization and functional analysis of *ilvA*, *ilvB* and *ilvC* genes from *Corynebacterium glutamicum*. *Gene.* 1992;112(1):113-116. DOI 10.1016/0378-1119(92)90311-c.
- Denina I., Paegle L., Prouza M., Holátko J., Pátek M., Nesvera J., Ruklisha M. Factors enhancing L-valine production by the growth-limited L-isoleucine auxotrophic strain *Corynebacterium glutamicum* DeltailvA DeltailvB ilvNM13 (pECKAilvBNC). *J. Ind. Microbiol. Biotechnol.* 2010;37(7):689-699. DOI 10.1007/s10295-010-0712-y.
- D'Este M., Alvarado-Morales M., Angelidaki I. Amino acids production focusing on fermentation technologies – A review. *Biotechnol. Adv.* 2017;36(1):14-25. DOI 10.1016/j.biotechadv.2017.09.001.
- Dimou A., Tsimihodimos V., Bairaktari E. The critical role of the branched chain amino acids (BCAAs) catabolism-regulating enzymes, branched-chain aminotransferase (BCAT) and branched-chain  $\alpha$ -keto acid dehydrogenase (BCKD), in human pathophysiology. *Int. J. Mol. Sci.* 2022;23(7):4022. DOI 10.3390/ijms23074022.
- Dostálová H., Holátko J., Busche T., Rucká L., Rapoport A., Halada P., Nešvera J., Kalinowski J., Pátek M. Assignment of sigma factors of RNA polymerase to promoters in *Corynebacterium glutamicum*. *AMB Express.* 2017;7(1):133. DOI 10.1186/s13568-017-0436-8.
- Dusch N., Pühler A., Kalinowski J. Expression of the *Corynebacterium glutamicum panD* gene encoding L-aspartate-alpha-decarboxylase leads to pantothenate overproduction in *Escherichia coli*. *Appl. Environ. Microbiol.* 1999;65(4):1530-1539. DOI 10.1128/AEM.65.4.1530-1539.1999.
- Ebbighausen H., Weil B., Krämer R. Transport of branched-chain amino acids in *Corynebacterium glutamicum*. *Arch. Microbiol.* 1989; 151(3):238-244. DOI 10.1007/BF00413136.
- Eggeling L., Cordes C., Eggeling L., Sahn H. Regulation of aceto-hydroxy acid synthase in *Corynebacterium glutamicum* during fermentation of alpha-ketobutyrate to L-isoleucine. *Appl. Microbiol. Biotechnol.* 1987;25(4):346-351. DOI 10.1007/BF00252545.
- Eggeling L. Exporters for production of amino acids and other small molecules. *Adv. Biochem. Eng. Biotechnol.* 2016;159:199-225. DOI 10.1007/10\_2016\_32.
- Eggeling L., Sahn H. New ubiquitous translocators: amino acid export by *Corynebacterium glutamicum* and *Escherichia coli*. *Arch. Microbiol.* 2003;180(3):155-160. DOI 10.1007/s00203-003-0581-0.
- Eikmanns B., Blombach B. The pyruvate dehydrogenase complex of *Corynebacterium glutamicum*: an attractive target for metabolic engineering. *J. Biotechnol.* 2014;192(Pt. B):339-345. DOI 10.1016/j.jbiotec.2013.12.019.
- Elišáková V., Pátek M., Holátko J., Nesvera J.N., Leyval D., Goergen J.L., Delaunay S. Feedback-resistant aceto-hydroxy acid synthase increases valine production in *Corynebacterium glutami-*

- cum. Appl. Environ. Microbiol.* 2005;71(1):207-213. DOI 10.1128/aem.71.1.207-213.2005.
- Engels V., Wendisch V.F. The DeoR-type regulator SugR represses expression of *ptsG* in *Corynebacterium glutamicum*. *J. Bacteriol.* 2007;189(8):2955-2966. DOI 10.1128/JB.01596-06.
- Goldbeck O., Eck A.W., Seibold G.M. Real time monitoring of NADPH concentrations in *Corynebacterium glutamicum* and *Escherichia coli* via the genetically encoded sensor mBFP. *Front. Microbiol.* 2018;9:2564. DOI 10.3389/fmicb.2018.02564.
- Guo Y., Han M., Xu J., Zhang W. Analysis of acetohydroxyacid synthase variants from branched-chain amino acids-producing strains and their effects on the synthesis of branched-chain amino acids in *Corynebacterium glutamicum*. *Protein Expr. Purif.* 2015;109:106-112. DOI 10.1016/j.pep.2015.02.006.
- Guo Y., Han M., Yan W., Xu J., Zhang W. Generation of branched-chain amino acids resistant *Corynebacterium glutamicum* acetohydroxy acid synthase by site-directed mutagenesis. *Biotechnol. Bioproc. Eng.* 2014;19:456-467. DOI 10.1007/s12257-013-0843-x.
- Han G., Xu N., Sun X., Chen J., Chen C., Wang Q. Improvement of L-valine production by atmospheric and room temperature plasma mutagenesis and high-throughput screening in *Corynebacterium glutamicum*. *ACS Omega.* 2020;5(10):4751-4758. DOI 10.1021/acsomega.9b02747.
- Harst A., Albaum S.P., Bojarzyn T., Trötschel C., Poetsch A. Proteomics of FACS-sorted heterogeneous *Corynebacterium glutamicum* populations. *J. Proteomics.* 2017;160:1-7. DOI 10.1016/j.jprot.2017.03.01.
- Hasegawa S., Suda M., Uematsu K., Natsuma Y., Hiraga K., Jojima T., Inui M., Yukawa H. Engineering of *Corynebacterium glutamicum* for high-yield L-valine production under oxygen deprivation conditions. *Appl. Environ. Microbiol.* 2013;79(4):1250-1257. DOI 10.1128/aem.02806-12.
- Hasegawa S., Uematsu K., Natsuma Y., Suda M., Hiraga K., Jojima T., Inui M., Yukawa H. Improvement of the redox balance increases L-valine production by *Corynebacterium glutamicum* under oxygen deprivation conditions. *Appl. Environ. Microbiol.* 2012;78(3):865-875. DOI 10.1128/aem.07056-11.
- Hemmerich J., Tenhaef N., Steffens C., Kappelmann J., Weiske M., Reich S.J., Wiechert W., Oldiges M., Noack S. Less sacrifice, more insight: Repeated low-volume sampling of microbioreactor cultivations enables accelerated deep phenotyping of microbial strain libraries. *Biotechnol. J.* 2018;14(9):e1800428. DOI 10.1002/biot.201800428.
- Hermann T., Kramer R. Mechanism and regulation of isoleucine excretion in *Corynebacterium glutamicum*. *Appl. Environ. Microbiol.* 1996;62(9):3238-3244. DOI 10.1128/aem.62.9.3238-3244.1996.
- Holátko J., Elišáková V., Prouza M., Sobotka M., Nesvera J., Patek M. Metabolic engineering of the L-valine biosynthesis pathway in *Corynebacterium glutamicum* using promoter activity modulation. *J. Biotechnol.* 2009;139(3):203-210. DOI 10.1016/j.jbiotec.2008.12.005.
- Holeček M. Branched-chain amino acids in health and disease: metabolism, alterations in blood plasma, and as supplements. *Nutr. Metab. (Lond).* 2018;15:33. DOI 10.1186/s12986-018-0271-1.
- Holen J.P., Tokach M.D., Woodworth J.C., DeRouchey J.M., Gebhardt J.T., Titgemeyer E.C., Goodband R.D. A review of branched-chain amino acids in lactation diets on sow and litter growth performance. *Transl. Anim. Sci.* 2022;6(1):txac017. DOI 10.1093/tas/txac017.
- Hou X.H., Chen X.D., Zhang Y., Qian H., Zhang W.G. L-valine production with minimization of by-products' synthesis in *Corynebacterium glutamicum* and *Brevibacterium flavum*. *Amino Acids.* 2012a;43(6):2301-2311. DOI 10.1007/s00726-012-1308-9.
- Hou X.H., Ge X.Y., Wu D., Qian H., Zhang W.G. Improvement of L-valine production at high temperature in *Brevibacterium flavum* by overexpressing *ilvEBN(r)C* genes. *J. Ind. Microbiol. Biotechnol.* 2012b;39(1):63-72. DOI 10.1007/s10295-011-1000-1.
- Jian H., Miao S., Liu Y., Li H., Zhou W., Wang X., Dong X., Zou X. Effects of dietary valine levels on production performance, egg quality, antioxidant capacity, immunity, and intestinal amino acid absorption of laying hens during the peak lay period. *Animals (Basel).* 2021;11(7):1972. DOI 10.3390/ani11071972.
- Jiang L.Y., Zhang Y.Y., Li Z., Liu J.Z. Metabolic engineering of *Corynebacterium glutamicum* for increasing the production of L-ornithine by increasing NADPH availability. *J. Ind. Microbiol. Biotechnol.* 2013;40(10):1143-1151. DOI 10.1007/s10295-013-1306-2.
- Jojima T., Fujii M., Mori E., Inui M., Yukawa H. Engineering of sugar metabolism of *Corynebacterium glutamicum* for production of amino acid L-alanine under oxygen deprivation. *Appl. Microbiol. Biotechnol.* 2010;87(1):159-165. DOI 10.1007/s00253-010-2493-7.
- Jojima T., Noburyu R., Sasaki M., Tajima T., Suda M., Yukawa H., Inui M. Metabolic engineering for improved production of ethanol by *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* 2015;99(3):1165-1172. DOI 10.1007/s00253-014-6223-4.
- Jones C.M., Hernandez Lozada N.J., Pfleger B.F. Efflux systems in bacteria and their metabolic engineering applications. *Appl. Microbiol. Biotechnol.* 2015;99(22):9381-9393. DOI 10.1007/s00253-015-6963-9.
- Kabus A., Georgi T., Wendisch V.F., Bott M. Expression of the *Escherichia coli pntAB* genes encoding a membrane-bound transhydrogenase in *Corynebacterium glutamicum* improves L-lysine formation. *Appl. Microbiol. Biotechnol.* 2007;75(1):47-53. DOI 10.1007/s00253-006-0804-9.
- Kainulainen H., Hulmi J.J., Kujala U.M. Potential role of branched-chain amino acid catabolism in regulating fat oxidation. *Exerc. Sport Sci. Rev.* 2013;41(4):194-200. DOI 10.1097/JES.0b013e3182a4e6b6.
- Kang K.Y., Kim M.S., Lee M.S., Oh J.J., An S., Park D., Heo I.K., Lee H.K., Song S.W., Kim S.D. Genotoxicity and acute toxicity evaluation of the three amino acid additives with *Corynebacterium glutamicum* biomass. *Toxicol. Rep.* 2020;7:241-253. DOI 10.1016/j.toxrep.2020.01.013.
- Karau A., Grayson I. Amino acids in human and animal nutrition. *Adv. Biochem. Eng. Biotechnol.* 2014;143:189-228. DOI 10.1007/10\_2014\_269.
- Kataoka N., Vangnai A.S., Pongtharangkul T., Yakushi T., Wada M., Yokota A., Matsushita K. Engineering of *Corynebacterium glutamicum* as a prototrophic pyruvate-producing strain: Characterization of a *ramA*-deficient mutant and its application for metabolic engineering. *Biosci. Biotechnol. Biochem.* 2019;83(2):372-380. DOI 10.1080/09168451.2018.1527211.
- Kawaguchi T., Izumi N., Charlton M.R., Sata M. Branched-chain amino acids as pharmacological nutrients in chronic liver disease. *Hepatology.* 2011;54(3):1063-1070. DOI 10.1002/hep.24412.
- Keilhauer C., Eggeling L., Sahm H. Isoleucine synthesis in *Corynebacterium glutamicum*: molecular analysis of the *ilvB-ilvN-ilvC* operon. *J. Bacteriol.* 1993;175(17):5595-5603. DOI 10.1128/jb.175.17.5595-5603.1993.
- Kennerknecht N., Sahm H., Yen M.R., Pátek M., Saier M.H. Jr., Eggeling L. Export of L-isoleucine from *Corynebacterium glutamicum*: a two-gene-encoded member of a new translocator family. *J. Bacteriol.* 2002;184(14):3947-3956. DOI 10.1128/jb.184.14.3947-3956.2002.
- Koduru L., Lakshmanan M., Lee D.Y. In silico model-guided identification of transcriptional regulator targets for efficient strain design. *Microb. Cell Fact.* 2018;17(1):167. DOI 10.1186/s12934-018-1015-7.
- Krause F.S., Blombach B., Eikmanns B.J. Metabolic engineering of *Corynebacterium glutamicum* for 2-ketoisovalerate production. *Appl. Environ. Microbiol.* 2010a;76(24):8053-8061. DOI 10.1128/aem.01710-10.
- Krause F.S., Henrich A., Blombach B., Kramer R., Eikmanns B.J., Seibold G.M. Increased glucose utilization in *Corynebacterium glutamicum* by use of maltose, and its application for the improvement of L-valine productivity. *Appl. Environ. Microbiol.* 2010b;76(1):370-374. DOI 10.1128/aem.01553-09.
- Lange C., Mustafi N., Frunzke J., Kennerknecht N., Wessel M., Bott M., Wendisch V.F. Lrp of *Corynebacterium glutamicum* controls expression of the *brnFE* operon encoding the export system

- for L-methionine and branched-chain amino acids. *J. Biotechnol.* 2012;158(4):231-241. DOI 10.1016/j.jbiotec.2011.06.003.
- Lange J., Münch E., Müller J., Busche T., Kalinowski J., Takors R., Blombach B. Deciphering the adaptation of *Corynebacterium glutamicum* in transition from aerobiosis via microaerobiosis to anaerobiosis. *Genes (Basel)*. 2018;9(6):297. DOI 10.3390/genes9060297.
- Lee D., Hong J., Kim K.J. Crystal structure and biochemical characterization of ketol-acid reductoisomerase from *Corynebacterium glutamicum*. *J. Agric. Food Chem.* 2019;67(31):8527-8535. DOI 10.1021/acs.jafc.9b03262.
- Leuchtenberger W., Huthmacher K., Drauz K. Biotechnological production of amino acids and derivatives: current status and prospects. *Appl. Microbiol. Biotechnol.* 2005;69(1):1-8. DOI 10.1007/s00253-005-0155-y.
- Leyval D., Uy D., Delaunay S., Goergen J.L., Engasser J.M. Characterisation of the enzyme activities involved in the valine biosynthetic pathway in a valine-producing strain of *Corynebacterium glutamicum*. *J. Biotechnol.* 2003;104(1-3):241-252. DOI 10.1016/s0168-1656(03)00162-7.
- Li N., Xu S., Du G., Chen J., Zhou J. Efficient production of L-homoserine in *Corynebacterium glutamicum* ATCC 13032 by redistribution of metabolic flux. *Biochem. Eng. J.* 2020a;161:107665. DOI 10.1016/j.bej.2020.107665.
- Li N., Zeng W., Xu S., Zhou J. Obtaining a series of native gradient promoter-5'-UTR sequences in *Corynebacterium glutamicum* ATCC 13032. *Microb. Cell. Fact.* 2020b;19(1):120. DOI 10.1186/s12934-020-01376-3.
- Li Y., Cong H., Liu B., Song J., Sun X., Zhang J., Yang Q. Metabolic engineering of *Corynebacterium glutamicum* for methionine production by removing feedback inhibition and increasing NADPH level. *Antonie Van Leeuwenhoek*. 2016;109(9):1185-1197. DOI 10.1007/s10482-016-0719-0.
- Lindner S.N., Petrov D.P., Hagmann C.T., Henrich A., Krämer R., Eikmanns B.J., Wendisch V.F., Seibold G.M. Phosphotransferase system-mediated glucose uptake is repressed in phosphoglucosyltransferase-deficient *Corynebacterium glutamicum* strains. *Appl. Environ. Microbiol.* 2013;79(8):2588-2595. DOI 10.1128/AEM.03231-12.
- Liu Y., Li Y., Wang X. Acetohydroxyacid synthases: evolution, structure, and function. *Appl. Microbiol. Biotechnol.* 2016;100(20):8633-8649. DOI 10.1007/s00253-016-7809-9.
- Liu Y., Wang X., Zhan J., Hu J. The 138<sup>th</sup> residue of acetohydroxyacid synthase in *Corynebacterium glutamicum* is important for the substrate binding specificity. *Enzyme Microb. Technol.* 2019;129:109357. DOI 10.1016/j.enzmictec.2019.06.001.
- Liu Y., Zhang C., Zhang Y., Jiang X., Liang Y., Wang H., Li Y., Sun G. Association between excessive dietary branched-chain amino acids intake and hypertension risk in chinese population. *Nutrients*. 2022;14(13):2582. DOI 10.3390/nu14132582.
- Ma Y., Chen Q., Cui Y., Du L., Shi T., Xu Q., Ma Q., Xie X., Chen N. Comparative genomic and genetic functional analysis of industrial L-leucine- and L-valine-producing *Corynebacterium glutamicum* strains. *J. Microbiol. Biotechnol.* 2018a;28(11):1916-1927. DOI 10.4014/jmb.1805.05013.
- Ma Y., Cui Y., Du L., Liu X., Xie X., Chen N. Identification and application of a growth-regulated promoter for improving L-valine production in *Corynebacterium glutamicum*. *Microb. Cell. Fact.* 2018b;17(1):185. DOI 10.1186/s12934-018-1031-7.
- Magnus J.B., Oldiges M., Takors R. The identification of enzyme targets for the optimization of a valine producing *Corynebacterium glutamicum* strain using a kinetic model. *Biotechnol. Prog.* 2009;25(3):754-762. DOI 10.1002/btpr.184.
- Marienhagen J., Eggeling L. Metabolic function of *Corynebacterium glutamicum* aminotransferases AlaT and AvtA and impact on L-valine production. *Appl. Environ. Microbiol.* 2008;74(24):7457-7462. DOI 10.1128/AEM.01025-08.
- Marienhagen J., Kennerknecht N., Sahn H., Eggeling L. Functional analysis of all aminotransferase proteins inferred from the genome sequence of *Corynebacterium glutamicum*. *J. Bacteriol.* 2005;187(22):7639-7646. DOI 10.1128/JB.187.22.7639-7646.2005.
- Marx A., Striegel K., de Graaf A.A., Sahn H., Eggeling L. Response of the central metabolism of *Corynebacterium glutamicum* to different flux burdens. *Biotechnol. Bioeng.* 1997;56(2):168-180. DOI 10.1002/(SICI)1097-0290(19971020)56:2<168::AID-BIT6>3.0.CO;2-N.
- Michel A., Koch-Koerfges A., Krumbach K., Brocker M., Bott M. Anaerobic growth of *Corynebacterium glutamicum* via mixed-acid fermentation. *Appl. Environ. Microbiol.* 2015;81(21):7496-7508. DOI 10.1128/AEM.02413-15.
- Möckel B., Eggeling L., Sahn H. Functional and structural analyses of threonine dehydratase from *Corynebacterium glutamicum*. *J. Bacteriol.* 1992;174(24):8065-8072. DOI 10.1128/jb.174.24.8065-8072.1992.
- Morbach S., Junger C., Sahn H., Eggeling L. Attenuation control of *ilvBNC* in *Corynebacterium glutamicum*: evidence of leader peptide formation without the presence of a ribosome binding site. *J. Biosci. Bioeng.* 2000;90(5):501-507. DOI 10.1016/S1389-1723(01)80030-X.
- Moritz B., Striegel K., De Graaf A.A., Sahn H. Kinetic properties of the glucose-6-phosphate and 6-phosphogluconate dehydrogenases from *Corynebacterium glutamicum* and their application for predicting pentose phosphate pathway flux *in vivo*. *Eur. J. Biochem.* 2000;267(12):3442-3452. DOI 10.1046/j.1432-1327.2000.01354.x.
- Okino S., Suda M., Fujikura K., Inui M., Yukawa H. Production of D-lactic acid by *Corynebacterium glutamicum* under oxygen deprivation. *Appl. Microbiol. Biotechnol.* 2008;78(3):449-454. DOI 10.1007/s00253-007-1336-7.
- Park J.H., Lee S.Y. Fermentative production of branched chain amino acids: a focus on metabolic engineering. *Appl. Microbiol. Biotechnol.* 2010;85(3):491-506. DOI 10.1007/s00253-009-2307-y.
- Pérez-García F., Jorge J.M.P., Dreyszas A., Risse J.M., Wendisch V.F. Efficient production of the dicarboxylic acid glutarate by *Corynebacterium glutamicum* via a novel synthetic pathway. *Front. Microbiol.* 2018;9:2589. DOI 10.3389/fmicb.2018.02589.
- Pérez-García F., Wendisch V.F. Transport and metabolic engineering of the cell factory *Corynebacterium glutamicum*. *FEMS Microbiol. Lett.* 2018;365(16):fny166. DOI 10.1093/femsle/fny166.
- Qin T., Hu X., Hu J., Wang X. Metabolic engineering of *Corynebacterium glutamicum* strain ATCC13032 to produce L-methionine. *Biotechnol. Appl. Biochem.* 2015;62(4):563-673. DOI 10.1002/bab.1290.
- Radmacher E., Vaitsikova A., Burger U., Krumbach K., Sahn H., Eggeling L. Linking central metabolism with increased pathway flux: L-valine accumulation by *Corynebacterium glutamicum*. *Appl. Environ. Microbiol.* 2002;68(5):2246-2250. DOI 10.1128/aem.68.5.2246-2250.2002.
- Ruklisha M., Paegle L., Denina I. L-Valine biosynthesis during batch and fed-batch cultivations of *Corynebacterium glutamicum*: Relationship between changes in bacterial growth rate and intracellular metabolism. *Proc. Biochem.* 2007;40(4):634-640. DOI 10.1016/j.procbio.2006.11.008.
- Sahn H., Eggeling L. D-pantothenate synthesis in *Corynebacterium glutamicum* and use of *panBC* and genes encoding L-valine synthesis for D-pantothenate overproduction. *Appl. Environ. Microbiol.* 1999;65(5):1973-1979. DOI 10.1128/AEM.65.5.1973-1979.1999.
- Savrasova E.A., Stoyanova N.V. Application of leucine dehydrogenase Bcd from *Bacillus subtilis* for L-valine synthesis in *Escherichia coli* under microaerobic conditions. *Heliyon*. 2019;5(4):e01406. DOI 10.1016/j.heliyon.2019.e01406.
- Schwentner A., Feith A., Münch E., Busche T., Rückert C., Kalinowski J., Takors R., Blombach B. Metabolic engineering to guide evolution – Creating a novel mode for L-valine production with *Corynebacterium glutamicum*. *Metab. Eng.* 2018;47:31-41. DOI 10.1016/j.ymben.2018.02.015.
- Shi F., Li K., Huan X., Wang X. Expression of NAD(H) kinase and glucose-6-phosphate dehydrogenase improve NADPH supply and L-isoleucine biosynthesis in *Corynebacterium glutamicum* ssp. *lactofermentum*. *Appl. Biochem. Biotechnol.* 2013;171(2):504-521. DOI 10.1007/s12010-013-0389-6.

- Shi F., Luan M., Li Y. Ribosomal binding site sequences and promoters for expressing glutamate decarboxylase and producing  $\gamma$ -aminobutyrate in *Corynebacterium glutamicum*. *AMB Express*. 2018; 8(1):61. DOI 10.1186/s13568-018-0595-2.
- Shou J., Chen P.J., Xiao W.H. The effects of BCAAs on insulin resistance in athletes. *J. Nutr. Sci. Vitaminol. (Tokyo)*. 2019;65(5):383-389. DOI 10.3177/jnsv.65.383.
- Siedler S., Lindner S.N., Bringer S., Wendisch V.F., Bott M. Reductive whole-cell biotransformation with *Corynebacterium glutamicum*: improvement of NADPH generation from glucose by a cyclized pentose phosphate pathway using *pfkA* and *gapA* deletion. *Appl. Microbiol. Biotechnol.* 2013;97(1):143-152. DOI 10.1007/s00253-012-4314-7.
- Tarutina M.G., Raevskaya N.M., Shustikova T.E., Ryabchenko L.E., Yanenko A.S. Assessment of effectiveness of *Corynebacterium glutamicum* promoters and their application for the enhancement of gene activity in lysine-producing bacteria. *Appl. Biochem. Microbiol.* 2016;52(7):692-698. DOI 10.1134/S0003683816070073.
- Tauch A., Hermann T., Burkovski A., Kramer R., Puhler A., Kalinowski J. Isoleucine uptake in *Corynebacterium glutamicum* ATCC 13032 is directed by the *brnQ* gene product. *Arch. Microbiol.* 1998;169(4):303-312. DOI 10.1007/s002030050576.
- Trotschel C., Deutenberg D., Bathe B., Burkovski A., Kramer R. Characterization of methionine export in *Corynebacterium glutamicum*. *J. Bacteriol.* 2005;187(11):3786-3794. DOI 10.1128/jb.187.11.3786-3794.2005.
- Vasicová P., Pátek M., Nesvera J., Sahm H., Eikmanns B. Analysis of the *Corynebacterium glutamicum* *dapA* promoter. *J. Bacteriol.* 1999; 181(19):6188-6191. DOI 10.1128/JB.181.19.6188-6191.1999.
- Vogt M., Haas S., Klaffl S., Polen T., Eggeling L., van Ooyen J., Bott M. Pushing product formation to its limit: metabolic engineering of *Corynebacterium glutamicum* for L-leucine overproduction. *Metab. Eng.* 2014;22:40-52. DOI 10.1016/j.ymben.2013.12.001.
- Wang X., Zhang H., Quinn P.J. Production of L-valine from metabolically engineered *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* 2018;102(10):4319-4330. DOI 10.1007/s00253-018-8952-2.
- Wang Y.Y., Shi K., Chen P., Zhang F., Xu J.Z., Zhang W.G. Rational modification of the carbon metabolism of *Corynebacterium glutamicum* to enhance L-leucine production. *J. Ind. Microbiol. Biotechnol.* 2020;47(6-7):485-495. DOI 10.1007/s10295-020-02282-8.
- Wang Y.Y., Xu J.Z., Zhang W.G. Metabolic engineering of L-leucine production in *Escherichia coli* and *Corynebacterium glutamicum*: a review. *Crit. Rev. Biotechnol.* 2019a;39(5):633-647. DOI 10.1080/07388551.2019.1577214.
- Wang Y.Y., Zhang F., Xu J.Z., Zhang W.G., Chen X.L., Liu L.M. Improvement of L-leucine production in *Corynebacterium glutamicum* by altering the redox flux. *Int. J. Mol. Sci.* 2019b;20(8):2020. DOI 10.3390/ijms20082020.
- Wang Z., Chen T., Ma X., Shen Z., Zhao X. Enhancement of riboflavin production with *Bacillus subtilis* by expression and site-directed mutagenesis of *zwf* and *gnd* gene from *Corynebacterium glutamicum*. *Bioresour. Technol.* 2011;102(4):3934-3940. DOI 10.1016/j.biortech.2010.11.120.
- Wei H., Ma Y., Chen Q., Cui Y., Du L., Ma Q., Li Y., Xie X., Chen N. Identification and application of a novel strong constitutive promoter in *Corynebacterium glutamicum*. *Ann. Microbiol.* 2018;68:375-382. DOI 10.1007/s13213-018-1344-0.
- Wieschalka S., Blombach B., Bott M., Eikmanns B.J. Bio-based production of organic acids with *Corynebacterium glutamicum*. *Microb. Biotechnol.* 2012;6(2):87-102. DOI 10.1111/1751-7915.12013.
- Xie X., Xu L., Shi J., Xu Q., Chen N. Effect of transport proteins on L-isoleucine production with the L-isoleucine-producing strain *Corynebacterium glutamicum* YILW. *J. Ind. Microbiol. Biotechnol.* 2012;39(10):1549-1556. DOI 10.1007/s10295-012-1155-4.
- Xu J., Han M., Zhang J., Guo Y., Zhang W. Metabolic engineering *Corynebacterium glutamicum* for the L-lysine production by increasing the flux into L-lysine biosynthetic pathway. *Amino Acids*. 2014;46(9):2165-2175. DOI 10.1007/s00726-014-1768-1.
- Xu J.Z., Yu H.B., Han M., Liu L.M., Zhang W.G. Metabolic engineering of glucose uptake systems in *Corynebacterium glutamicum* for improving the efficiency of L-lysine production. *J. Ind. Microbiol. Biotechnol.* 2019;46(7):937-949. DOI 10.1007/s10295-019-02170-w.
- Xu N., Wei L., Liu J. Recent advances in the applications of promoter engineering for the optimization of metabolite biosynthesis. *World J. Microbiol. Biotechnol.* 2019;35(2):33. DOI 10.1007/s11274-019-2606-0.
- Yamamoto K., Tsuchisaka A., Yukawa H. Branched-chain amino acids. *Adv. Biochem. Eng. Biotechnol.* 2017;159:103-128. DOI 10.1007/10\_2016\_28.
- Yamamoto S., Suda M., Niimi S., Inui M., Yukawa H. Strain optimization for efficient isobutanol production using *Corynebacterium glutamicum* under oxygen deprivation. *Biotechnol. Bioeng.* 2013; 110(11):2938-2948. DOI 10.1002/bit.24961.
- Yin L., Shi F., Hu X., Chen C., Wang X. Increasing L-isoleucine production in *Corynebacterium glutamicum* by overexpressing global regulator Lrp and two-component export system BrnFE. *J. Appl. Microbiol.* 2013;114(5):1369-1377. DOI 10.1111/jam.12141.
- Yin L., Zhao J., Chen C., Xu X., Wang X. Enhancing the carbon flux and NADPH supply to increase L-isoleucine production in *Corynebacterium glutamicum*. *Biotechnol. Bioproc. Eng.* 2014;19:132-142. DOI 10.1007/s12257-013-0416-z.
- Zhan M., Kan B., Dong J., Xu G., Han R., Ni Y. Metabolic engineering of *Corynebacterium glutamicum* for improved L-arginine synthesis by enhancing NADPH supply. *J. Ind. Microbiol. Biotechnol.* 2019;46(1):45-54. DOI 10.1007/s10295-018-2103-8.
- Zhang H., Li Y., Wang C., Wang X. Understanding the high L-valine production in *Corynebacterium glutamicum* VWB-1 using transcriptomics and proteomics. *Sci. Rep.* 2018;8(1):3632. DOI 10.1038/s41598-018-21926-5.
- Zhang J., Qian F., Dong F., Wang Q., Yang J., Jiang Y., Yang S. De novo engineering of *Corynebacterium glutamicum* for L-proline production. *ACS Synth. Biol.* 2020;9(7):1897-1906. DOI 10.1021/acssynbio.0c00249.
- Zhang S., Liu D., Mao Z., Mao Y., Ma H., Chen T., Zhao X., Wang Z. Model-based reconstruction of synthetic promoter library in *Corynebacterium glutamicum*. *Biotechnol. Lett.* 2018;40(5):819-827. DOI 10.1007/s10529-018-2539-y.
- Zhang Y., Liu Y., Zhang S., Ma W., Wang J., Yin L., Wang X. Metabolic engineering of *Corynebacterium glutamicum* WM001 to improve L-isoleucine production. *Biotechnol. Appl. Biochem.* 2021;68(3): 568-584. DOI 10.1002/bab.1963.
- Zheng L., Zuo F., Zhao S., He P., Wei H., Xiang Q., Pang J., Peng J. Dietary supplementation of branched-chain amino acids increases muscle net amino acid fluxes through elevating their substrate availability and intramuscular catabolism in young pigs. *Br. J. Nutr.* 2017;117(7):911-922. DOI 10.1017/S0007114517000757.

**Благодарности.** Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации (проекты № 075-15-2019-1659 и 075-15-2019-1662).

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 07.08.2022. После доработки 26.10.2022. Принята к публикации 26.10.2022.

Перевод на английский язык <https://vavilov.elpub.ru/jour>

## Стратификации и слоения в фазовых портретах моделей генных сетей

В.П. Голубятников<sup>1, 3</sup> ✉, А.А. Акиншин<sup>2</sup>, Н.Б. Аюпова<sup>1, 3</sup>, Л.С. Минушкина<sup>3</sup>

<sup>1</sup> Институт математики им. С.Л. Соболева Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>2</sup> Российский исследовательский институт Huawei, Санкт-Петербург, Россия

<sup>3</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

✉ golubyatn@yandex.ru

**Аннотация.** Периодические процессы функционирования широкого класса генных сетей с хорошей точностью описываются предельными циклами многомерных систем дифференциальных уравнений кинетического типа. Такие системы, часто называемые в литературе динамическими, состоят по схемам положительных и отрицательных связей между компонентами моделируемых сетей. Искомые функции в уравнениях описывают зависимость от времени концентраций этих компонент. При планировании вычислительных экспериментов с подобными математическими моделями полезно предварительно описать качественное поведение ансамблей траекторий соответствующих динамических систем, в частности оценить области максимального правдоподобия начальных данных, исследовать обратные задачи идентификации параметров, особые точки этих систем, локализовать в фазовых портретах положение циклов, в том числе предельных, стратифицировать фазовые портреты на подобласти с качественно различным поведением траекторий и т.п. Такой априорный геометрический анализ рассматриваемых моделей генных сетей полностью аналогичен хрестоматийному разделу начальных курсов математики «Исследование функций и построение графиков», в котором описываются методы наглядного представления поведения кривых, определяемых уравнениями. В настоящей статье в фазовых портретах динамических систем, моделирующих функционирование кольцевых генных сетей, конструируются двумерные поверхности, инвариантные относительно сдвигов вдоль траекторий, – ансамбли траекторий. Просматривается естественная аналогия с классической конструкцией аналитической механики – с поверхностями уровня интегралов движения (энергия, импульс и др.). Такие поверхности образуют слоения в фазовых портретах динамических систем гамильтоновой механики. В отличие от задач механики, для рассматриваемых нами моделей генных сетей слоения, образуемые инвариантными поверхностями, имеют особенности, все их слои содержат на своих границах предельные циклы. Описание фазовых портретов динамических систем в терминах их стратификаций и ансамблей их траекторий позволит строить более реалистичные модели генных сетей с использованием аппарата статистической физики и теории стохастических дифференциальных уравнений.

Ключевые слова: осцилляции; положительные и отрицательные связи; модели генных сетей; фазовые портреты; инвариантные области и поверхности; инвариантные слоения; отображение Пуанкаре; теорема Гробмана–Хартмана; теорема Фробениуса–Перрона.

**Для цитирования:** Голубятников В.П., Акиншин А.А., Аюпова Н.Б., Минушкина Л.С. Стратификации и слоения в фазовых портретах моделей генных сетей. *Вавиловский журнал генетики и селекции*. 2022;26(8):758-764. DOI 10.18699/VJGB-22-91

## Stratifications and foliations in phase portraits of gene network models

V.P. Golubyatnikov<sup>1, 3</sup> ✉, A.A. Akinshin<sup>2</sup>, N.B. Ayupova<sup>1, 3</sup>, L.S. Minushkina<sup>3</sup>

<sup>1</sup> Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>2</sup> Huawei Russian Research Institute, St. Petersburg, Russia

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

✉ golubyatn@yandex.ru

**Abstract.** Periodic processes of gene network functioning are described with good precision by periodic trajectories (limit cycles) of multidimensional systems of kinetic-type differential equations. In the literature, such systems are often called dynamical, they are composed according to schemes of positive and negative feedback between components of these networks. The variables in these equations describe concentrations of these components as functions of time. In the preparation of numerical experiments with such mathematical models, it is useful to start with studies of qualitative behavior of ensembles of trajectories of the corresponding dynamical systems, in particular, to estimate the highest likelihood domain of the initial data, to solve inverse problems of parameter identification, to list the equilibrium points and their characteristics, to localize cycles in the phase portraits, to construct stratification of the phase portraits to subdomains with different qualities of trajectory behavior, etc. Such an *à priori* geometric analysis of the dynamical systems is quite analogous to the basic section “Investigation of functions and plot of their graphs” of Calculus, where the methods of qualitative studies of

shapes of curves determined by equations are exposed. In the present paper, we construct ensembles of trajectories in phase portraits of some dynamical systems. These ensembles are 2-dimensional surfaces invariant with respect to shifts along the trajectories. This is analogous to classical construction in analytic mechanics, i.e. the level surfaces of motion integrals (energy, kinetic moment, etc.). Such surfaces compose foliations in phase portraits of dynamical systems of Hamiltonian mechanics. In contrast with this classical mechanical case, the foliations considered in this paper have singularities: all their leaves have a non-empty intersection, they contain limit cycles on their boundaries. Description of the phase portraits of these systems at the level of their stratifications, and that of ensembles of trajectories allows one to construct more realistic gene network models on the basis of methods of statistical physics and the theory of stochastic differential equations.

Key words: oscillations; positive and negative feedbacks; gene network models; phase portraits; invariant domains and surfaces; invariant foliations; Poincaré map; Grobman–Hartman theorem; Frobenius–Peron theorem.

**For citation:** Golubyatnikov V.P., Akinshin A.A., Ayupova N.B., Minushkina L.S. Stratifications and foliations in phase portraits of gene network models. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):758-764. DOI 10.18699/VJGB-22-91

## Введение

Исследования вопросов существования периодических траекторий (циклов) в фазовых портретах систем нелинейных дифференциальных уравнений, моделирующих динамику разнообразных природных процессов, в настоящее время проводятся во многих областях прикладной математики. Выявление таких циклов, локализация их в фазовых портретах и описание характеристик, таких как устойчивость, (не)единственность и т. п., имеют давнюю историю (Poincaré, 1892). Такие задачи породили целый ряд направлений в чистой математике: качественную теорию дифференциальных уравнений, теорию динамических систем и др., оказавших, в свою очередь, заметное влияние на смежные с этими направлениями прикладные дисциплины. На их стыке возникли известные 16-я проблема Гильберта и проблема «центр-фокус», относящиеся к, казалось бы, совсем наглядному случаю двух уравнений с двумя неизвестными функциями одной переменной (времени).

В настоящей работе мы изучаем системы кинетических уравнений больших размерностей, моделирующие функционирование кольцевых генных сетей

$$\frac{dx_j}{dt} = f_j(x_{j-1}) - k_j x_j. \quad (1)$$

Здесь и ниже предполагается, что  $j = 1, 2, \dots, n; n \geq 3$  и что  $j-1 = n$ , если  $j = 1$ . Неотрицательные функции  $x_j(t)$  в этих уравнениях обозначают концентрации веществ, принадлежащих моделируемой генной сети, положительные коэффициенты  $k_j$  характеризуют скорости их деградации (Likhoshvai et al., 2020).

Если записать систему уравнений (1) в векторной форме:  $\frac{dX}{dt} = F(X)$ , где вектор-функция  $X(t)$  определяется своими координатными функциями  $x_j(t)$ , то дивергенция векторного поля  $F(X)$  оказывается постоянной и отрицательной:

$$\operatorname{div} F(X) \equiv -k_1 - k_2 - \dots - k_n < 0.$$

Хорошо известно (Арнольд, 2003), что в этом случае  $n$ -мерный объем любой конечной области фазового портрета уменьшается экспоненциально с ростом  $t$  при сдвигах точек этой области вдоль траекторий системы (1). Это вовсе не означает, что каждая такая конечная область сжимается в точку. Для рассматриваемых в настоящей работе динамических систем такими предельными множествами

оказываются двумерные инвариантные поверхности в их  $n$ -мерных фазовых портретах.

Будем называть динамическую систему (1) блочно-линейной, если для всех  $j$  каждая функция  $f_j$ , описывающая скорость синтеза  $j$ -й компоненты генной сети, является ступенчатой (пороговой)

$$f_j(y) \equiv L_j(y) = k_j a_j, \text{ если } 0 \leq y \leq 1; L_j(y) \equiv 0, \text{ если } y > 1; \\ \text{либо } f_j(y) \equiv \Gamma_j(y) = 0, \text{ если } 0 \leq y \leq 1; \Gamma_j(y) \equiv k_j a_j, \text{ если } y > 1.$$

Здесь  $a_j$  – некоторые положительные постоянные. Убывающие функции  $L_j$  описывают отрицательные связи в генной сети, а возрастающие функции  $\Gamma_j$  соответствуют положительным связям.

Для частного случая  $k_j = 1$  при всех  $j$  исследование циклов подобных систем проводилось в работах (Glass, Pasternack, 1978; Акиншин и др., 2013; Аюпова, Голубятников, 2014; Голубятников, Градов, 2020). При этом же ограничении вопросы существования циклов изучались для ряда гладких аналогов таких динамических систем в (Elowitz, Leibler, 2000; Глызин и др., 2016; Колесов и др., 2016) в случаях, когда эти системы симметричны относительно циклических перестановок пар переменных  $x_j$ .

В недавних публикациях (Голубятников, Иванов, 2018; Голубятников, Минушкина, 2019; Golubyatnikov, Minushkina, 2020; Likhoshvai et al., 2020; Иванов, 2022) существование, единственность и устойчивость циклов у блочно-линейных динамических систем различных размерностей изучались при произвольных положительных коэффициентах  $k_j$  с помощью стратификации фазовых портретов на подобласти в зависимости от поведения траекторий. Показано, что у таких систем циклы существуют в том и только в том случае, когда  $a_j > 1$  при всех  $j$ , и что параллелепипед  $Q^n = [0, a_1] \times [0, a_2] \times \dots \times [0, a_n]$ , лежащий в положительном октанте пространства  $\mathbf{R}^n$ , является положительно инвариантной областью динамической системы (1). Это значит, что траектории всех точек области  $Q^n$  из нее не выходят и что все циклы системы (1) лежат во внутренности  $Q^n$ . В дальнейшем мы будем рассматривать только динамические системы вида (1), у которых  $a_j > 1$  при всех  $j$ . Физическая интерпретация этого условия состоит в том, что для каждой компоненты моделируемой генной сети максимальная скорость ее синтеза больше показателя скорости ее деградации.

Область  $Q^n$  разбивается плоскостями  $x_j = 1$  на  $2^n$  более мелких параллелепипедов, которые мы будем называть блоками и нумеровать бинарными мультииндексами:  $\{\varepsilon_1 \varepsilon_2 \dots \varepsilon_n\} = I_1(\varepsilon_1) \times I_2(\varepsilon_2) \times \dots \times I_n(\varepsilon_n)$ . Здесь каждый индекс  $\varepsilon_j$  принимает значение 0 или 1, и  $I_j(0) = [0, 1]$ ,  $I_j(1) = (1, a_j]$ . Обозначим общую точку всех этих блоков через  $E$  (все ее координаты равны единице). В каждом из таких блоков уравнения системы (1) принимают простейший линейный вид

$$\frac{dx_j}{dt} = k_j(x_j - a_j(1 - \varepsilon_{j-1})),$$

и решение задачи Коши для этой системы имеет несложное представление

$$x_j(t) = a_j(1 - \varepsilon_{j-1}) + (x_j(0) - a_j(1 - \varepsilon_{j-1})) \exp(-k_j t). \quad (2)$$

В этой работе для блочно-линейных динамических систем малых размерностей, моделирующих функционирование генных сетей, мы изучаем поведение ансамблей их траекторий и устанавливаем существование семейств двумерных поверхностей, инвариантных относительно сдвигов вдоль траекторий данных систем и содержащих их предельные циклы. Это позволяет значительно упрощать качественный анализ поведения траекторий этих систем и интерпретацию вычислительных экспериментов с такими моделями генных сетей.

### Трехмерная динамическая система

В работах (Голубятников и др., 2018; Голубятников, Иванов, 2018) рассматривалась блочно-линейная динамическая система размерности 3

$$\frac{dx_1}{dt} = L_1(x_3) - k_1 x_1; \quad \frac{dx_2}{dt} = L_2(x_1) - k_2 x_2; \quad \frac{dx_3}{dt} = L_3(x_2) - k_3 x_3. \quad (3)$$

Траектории всех точек блока  $\{001\}$  проходят через шесть блоков разбиения области  $Q^3$  из блока в блок только в соответствии со стрелками диаграммы

$$\dots \rightarrow \{001\} \rightarrow \{011\} \rightarrow \{010\} \rightarrow \{110\} \rightarrow \{100\} \rightarrow \{101\} \rightarrow \{001\} \dots \quad (4)$$

Обозначим через  $W_1^3$  объединение перечисленных в диаграмме блоков – это тоже положительно инвариантная область системы (3), ее внутренность гомеоморфна тору. Отметим, что траектории точек двух блоков,  $\{000\}$  и  $\{111\}$ , со временем выходят из них в инвариантную область  $W_1^3$  и далее остаются в ней. Поэтому циклы системы (3) по этим двум блокам не проходят (Голубятников и др., 2018). Стратификация фазового портрета системы (3) состоит из двух частей – области  $W_1^3$  и объединения блоков  $\{000\}$  и  $\{111\}$ .

Рассмотрим двумерную грань  $F_0 = \{001\} \cap \{011\}$ , разделяющую блоки  $\{001\}$  и  $\{011\}$ , и остальные грани  $F_m$ , разделяющие соседние блоки диаграммы (4):

$$F_1 = \{011\} \cap \{010\}, F_2 = \{010\} \cap \{110\}, \\ F_3 = \{110\} \cap \{100\}, \dots, F_5 = \{101\} \cap \{001\}.$$

После прохода по всем шести стрелкам этой диаграммы траектории точек грани  $F_0$  возвращаются в нее, каждая траектория за свое время. Композиция  $\Psi: F_0 \rightarrow F_0$  всех шести переходов с граней  $F_m$  на грани  $F_{m+1}$ ,  $m = 0, 1, 2, 3, 4$ , и  $F_5 \rightarrow F_0$  называется отображением Пуанкаре.

Введем на грани  $F_0$  систему координат  $(w_1; w_2)$  с началом в точке  $E_3 = (1; 1; 1)$ , такую, что у всех точек этой грани

координаты  $w_1, w_2$  неотрицательны:  $w_1 = 1 - x_2; w_2 = x_3 - 1$ . Пусть отображение Пуанкаре представлено в этих координатах уравнением  $\Psi(w_1; w_2) = (\psi_1(w_1; w_2); \psi_2(w_1; w_2))$ . Основным техническим результатом работ (Голубятников и др., 2018; Голубятников, Иванов, 2018) была

**Лемма 1:** а) отображение Пуанкаре монотонно, если для точек  $A(v_1; v_2)$  и  $B(w_1; w_2)$  выполнены соотношения  $v_1 < w_1$  и  $v_2 < w_2$ , то  $\psi_1(v_1; v_2) < \psi_1(w_1; w_2)$  и  $\psi_2(v_1; v_2) < \psi_2(w_1; w_2)$ . Для такого отношения частичного порядка будем использовать обозначения:  $A < B, \Psi(A) < \Psi(B)$ ; б) если  $w_1$  и  $w_2$  достаточно малы, то  $w_1 < \psi_1(w_1; w_2)$  и  $w_2 < \psi_2(w_1; w_2)$ , т. е.  $B < \Psi(B)$ ; в) во всех точках грани  $F_0$  первые производные координатных функций  $\psi_1$  и  $\psi_2$  строго положительны, а все вторые производные этих функций строго отрицательны.

Это позволило установить, что отображение Пуанкаре  $\Psi: F_0 \rightarrow F_0$  имеет в точности две неподвижные точки; одна из них – точка  $E_3$ , лежащая на границе  $F_0$ , а другая, обозначим ее через  $P_*$ , лежит во внутренности грани  $F_0$  (Голубятников, Иванов, 2018). Траектория точки  $P_*$  после прохождения по блокам диаграммы (4) возвращается в эту точку и, таким образом, является циклом. Поскольку отображение  $\Psi$  имеет всего одну нетривиальную неподвижную точку,  $P_*$ , других циклов у системы (3) нет.

В той же работе для неподвижных точек отображения Пуанкаре,  $E_3$  и  $P_*$ , были вычислены матрицы Якоби  $J_2(E_3)$ ,  $J_2(P_*)$  и показано, что у матрицы  $J_2(P_*)$  собственные числа  $\lambda_1(P_*), \lambda_2(P_*)$  различны, положительны и не превосходят единицу, что означает экспоненциальную устойчивость цикла системы (3). Будем обозначать этот цикл, обнаруженный в работе (Голубятников и др., 2018), через  $C_3$ . Из леммы 1 также следует, что обе эти матрицы Якоби положительны и к ним можно применять теорему Фробениуса–Перрона (Гантмахер, 1967).

Отметим, что у матрицы Якоби  $J_2(E_3)$  определитель равен единице, для ее собственных чисел  $\lambda_1(E_3), \lambda_2(E_3)$  выполнены соотношения  $\lambda_1(E_3) > 1 > \lambda_2(E_3) > 0$ , и поэтому для отображения  $\Psi$  выполняются условия теоремы Гробмана–Хартмана (Хартман, 1970), из которой следует, что в достаточно малой окрестности  $U(E_3) \subset F_0$  точки  $E_3$  отображение Пуанкаре линеаризуется с помощью некоторой непрерывной (вообще говоря, негладкой) замены переменных  $(w_1; w_2) \Rightarrow (u_1; u_2)$ . В такой системе координат оно имеет вид  $\Psi(u_1; u_2) = (\lambda_1(E_3) \cdot u_1; \lambda_2(E_3) \cdot u_2)$ .

Для достаточно малого  $\varepsilon > 0$  обозначим через  $T_\varepsilon^2$  содержащийся в  $U(E_3)$  треугольник  $0 \leq u_1 + u_2 < \varepsilon$  с вершиной в точке  $E_3$  и через  $\hat{F}_0$  – усеченную грань  $F_0 \cap T_\varepsilon^2$ .

Выберем в этой окрестности на оси  $E_3 u_1$  отрезки  $[0, \alpha_1]$  и  $[0, \alpha_0] \subset [0, \alpha_1]$ , где  $\alpha_1 = \lambda_1(E_3) \cdot \alpha_0$ . Пусть  $N_1$  и  $N_0$  соответственно – правые концы этих отрезков, тогда  $\Psi([0, \alpha_0]) = [0, \alpha_1]$  и  $\Psi(N_0) = N_1$ ; в исходной системе координат  $(w_1; w_2)$  отрезки  $[0, \alpha_0]$  и  $[0, \alpha_1]$  представлены дугами  $D_0 \subset D_1$  с общим концом в точке  $E_3$ . Рассмотрим действие итераций отображения Пуанкаре на эти дуги:

$$\Psi(D_0) = D_1 \subset D_2 = \Psi(D_1) \subset D_3 = \Psi(D_2) \subset D_4 \dots$$

Объединение  $D_*$  бесконечной последовательности вложенных друг в друга дуг  $D_k$  является непрерывной

монотонной дугой, соединяющей точки  $E_3$  и  $P_*$ , траектории точек дуги  $D_*$  после прохода по диаграмме (4) возвращаются на эту дугу: полуинтервал  $D_1 \setminus D_0$  переходит в полуинтервал  $D_2 \setminus D_1$ , который в свою очередь переходит в  $D_3 \setminus D_2$  и т. д. Таким образом, траектории точек дуги  $D_*$  образуют в инвариантной области  $W_1^3 \subset Q^3$  инвариантную (негладкую) поверхность  $\Sigma^2$ , краем которой является цикл  $C_3$ . По построению эта поверхность проходит через точку  $E_3$ .

Начиная такие построения малых отрезков  $[N_0, N_1]$  в окрестности  $U(E_3)$  с точек  $N_0$ , не лежащих на оси  $E_3 u_1$ , и рассматривая образы таких отрезков при итерациях отображения Пуанкаре  $\Psi$ , мы получаем семейство непрерывных монотонных дуг, выходящих из окрестности  $U(E_3)$ , и не содержащих точку  $E_3$ . Для каждой пары точек  $N_0, N_1 \subset U(E_3) \setminus E_3 u_1$  таких, что  $\Psi(N_0) = N_1$ , последовательность  $N_k = \Psi(N_{k-1})$  монотонно стремится к неподвижной точке  $P_*$  отображения Пуанкаре  $\Psi$  (Голубятников и др., 2018). При этом каждый отрезок  $[N_0, N_1]$  порождает, как и выше, монотонную дугу  $D_*(N_0)$ , инвариантную относительно отображения Пуанкаре; траектории точек такой дуги в свою очередь образуют инвариантную двумерную поверхность  $\Sigma^2(N_0)$ , пересекающуюся с поверхностью  $\Sigma^2$  в точности по циклу  $C_3$ .

Подобным же образом в области  $W_1^3$  строятся и инвариантные поверхности, не проходящие через окрестность  $U(E_3)$ . Пусть  $U(P_*) \subset \widehat{F}_0$  – окрестность нетривиальной неподвижной точки  $P_*$ , в которой отображение  $\Psi$  линеаризуется. Сохраним для таких линеаризующих координат обозначения  $(u_1; u_2)$ . Для достаточно малого  $\varepsilon > 0$  отображение Пуанкаре переводит эллипс  $S_1^1 \subset U(P_*)$  с уравнением  $\lambda_1(P_*)u_1^2 + \lambda_2(P_*)u_2^2 = \varepsilon^2$  в окружность  $S_0^1$  с уравнением  $u_1^2 + u_2^2 = \varepsilon^2$ . Соединим каждую точку  $M_0 \in S_0^1$  с точкой  $M_1 \in S_1^1$ , такой, что  $M_0 = \Psi(M_1)$ , отрезком  $I_1(M_0)$ . Все такие отрезки лежат в  $U(P_*)$  в кольце между  $S_0^1$  и  $S_1^1$ , и для каждого из этих отрезков можно построить последовательность инвариантных относительно отображения Пуанкаре непрерывных дуг  $D_k(M_0)$ , таких, что  $\Psi(D_k(M_0)) = D_{k-1}(M_0)$ . Для каждой такой дуги траектории ее точек образуют в  $W_1^3$  инвариантную поверхность, содержащую на своей границе цикл  $C_3$ .

**Теорема 1.** *В инвариантной области  $W_1^3$  динамической системы (3) существует двумерное инвариантное слоение, слои которого заполняют  $W_1^3$  и содержат на своих границах цикл  $C_3$ . Один из этих слоев проходит через точку  $E_3$ .*

#### Четырехмерная динамическая система

В работах (Аюпова, Голубятников 2019; Golubyatnikov, Minushkina, 2021) рассматривалась блочно-линейная динамическая система размерности 4

$$\frac{dx_1}{dt} = L_1(x_4) - k_1 x_1; \quad \frac{dx_r}{dt} = \Gamma_r(x_{r-1}) - k_r x_r; \quad r = 2, 3, 4. \quad (5)$$

В частном случае  $k_j = 1$  при всех  $j$  вопросы существования, единственности и устойчивости циклов у таких систем исследованы в работе (Glass, Pasternack, 1978). Гладкие аналоги подобных систем изучены в (Hastings et al., 1977; Mallet-Paret, Smith, 1990).

Инвариантная область  $Q^4$  системы (5) разбивается плоскостями  $x_j = 1$  на 16 блоков  $\{\varepsilon_1 \varepsilon_2 \varepsilon_3 \varepsilon_4\}$ . Перечисленные в следующей кольцевой диаграмме блоки этого разбиения образуют инвариантную область  $W_1^4$  в фазовом портрете системы (5)

$$\dots \rightarrow \{1111\} \rightarrow \{0111\} \rightarrow \{0011\} \rightarrow \{0001\} \rightarrow \{0000\} \rightarrow \{1000\} \rightarrow \{1100\} \rightarrow \{1110\} \rightarrow \{1111\} \rightarrow \dots \quad (6)$$

Стрелки этой диаграммы показывают единственные возможные направления переходов траекторий этой динамической системы из блока в блок. Область  $W_1^4$  – одна из двух частей стратификации фазового портрета системы (5). Для каждого из не перечисленных здесь блоков траектории его точек могут переходить в три соседних с ним блока, два из которых лежат в  $W_1^4$ , а один – в  $Q^4 \setminus W_1^4$  (это вторая часть указанной стратификации). Алгоритмы построения таких диаграмм для систем произвольных размерностей, гладких и блочно-линейных, описаны в работах (Казанцев, 2015; Кириллова, Минушкина, 2019).

Так же, как и выше, обозначим через  $F_0$  пересечение двух соседних блоков  $\{1111\} \cap \{0111\}$  в предыдущей диаграмме. Точки этой трехмерной грани после восьми шагов по стрелкам диаграммы (6) при сдвигах вдоль своих траекторий попадают на эту же грань  $F_0$ . Пусть  $\Psi_4: F_0 \rightarrow F_0$  – соответствующее отображение Пуанкаре,  $T_\varepsilon^3 \subset U(E_4)$  пирамида  $0 \leq u_1 + u_2 + u_3 < \varepsilon$  с вершиной в точке  $E$  и  $\widehat{F}_0$  – усеченная грань  $F_0 \setminus T_\varepsilon^3$ .

В работе (Golubyatnikov, Minushkina, 2021) показано, что для отображения  $\Psi_4$  справедливы все утверждения леммы 1. Откуда следует, что это отображение имеет в точности две неподвижные точки,  $E_4 = (1; 1; 1; 1)$  и  $P_*$ , лежащую во внутренности грани  $F_0$ . Это значит, что инвариантная область  $W_1^4$  системы (5) содержит в точности один цикл. Обозначим его через  $C_4$ . Там же были установлены следующие результаты:

- Лемма 2:** а) матрицы Якоби  $J_3(E_4)$ ,  $J_3(P_*)$  и их определители положительны;
- б)  $\det J_3(E_4) = \lambda_1(E_4) \cdot \lambda_2(E_4) \cdot \lambda_3(E_4) = 1$ ;
- в) модули собственных чисел матрицы  $J_3(P_*)$  меньше единицы.

Отсюда вытекает экспоненциальная устойчивость цикла  $C_4$  и возможность линеаризовать отображение Пуанкаре  $\Psi_4$  в малой окрестности  $U(P_*)$  его неподвижной точки  $P_*$ . Согласно теореме Фробениуса–Перрона, одно из собственных чисел матрицы  $J_3(P_*)$  положительно и превосходит модули остальных ее собственных чисел. То же относится и к собственным числам матрицы  $J_3(E_4)$ . Упорядочим собственные числа матриц Якоби в порядке убывания их модулей:  $\lambda_1 > |\lambda_2| \geq |\lambda_3|$ . Пусть  $(u_1; u_2; u_3)$  – координаты, в которых  $\Psi_4$  линейно

$$\Phi(u_1; u_2; u_3) = (\lambda_1(P_*) \cdot u_1; \lambda_2(P_*) \cdot u_2; \lambda_3(P_*) \cdot u_3).$$

Как и в случае системы (3), при достаточно малом  $\varepsilon > 0$  отображение Пуанкаре переводит эллипсоид  $S_1^2$  с уравнением  $\lambda_1(P_*)u_1^2 + |\lambda_2(P_*)|u_2^2 + |\lambda_3(P_*)|u_3^2 = \varepsilon^2$  в сферу  $S_0^2$  с уравнением  $u_1^2 + u_2^2 + u_3^2 = \varepsilon^2$ .

**Теорема 2.** *Если  $a_j > 1$  для всех  $j = 1, 2, 3, 4$ , и матрица Якоби  $J_3(E_4)$  отображения Пуанкаре не имеют собственных чисел, модуль которых равен единице, то*

в области  $W_1^4$  существует двумерное инвариантное слоение, слои которого заполняют всю эту область и содержат на своих границах цикл  $C_4$ . Один из этих слоев проходит через точку  $E_4$ .

### Динамические системы старших размерностей

В работах (Gaidov, Golubyatnikov, 2014; Аюпова, Голубятников, 2021) рассматривалась блочно-линейная динамическая система размерности 5

$$\dot{x}_1 = L_1(x_5) - k_1 x_1; \dot{x}_2 = L_2(x_1) - k_2 x_2; \dots \dot{x}_5 = L_5(x_4) - k_5 x_5, \quad (7)$$

для которой, как и выше, были построены инвариантная область  $Q^5 = [0, a_1] \times [0, a_2] \times \dots \times [0, a_5]$  и ее разбиение на блоки плоскостями  $x_j = 1$ . Десять блоков этого разбиения образуют страт  $W_1^5 \subset Q^5$ , инвариантный относительно сдвигов вдоль траекторий системы (7), которые переходят из блока в блок в соответствии со стрелками кольцевой диаграммы, аналогичной (4) и (6),

$$\begin{aligned} \dots \rightarrow \{10101\} \rightarrow \{00101\} \rightarrow \{01101\} \rightarrow \{01001\} \rightarrow \\ \{01011\} \rightarrow \{01010\} \rightarrow \{11010\} \rightarrow \{10010\} \rightarrow \\ \{10110\} \rightarrow \{10100\} \rightarrow \{10101\} \rightarrow \dots \end{aligned}$$

Точки четырехмерной грани  $F_0^4 = \{10101\} \cap \{00101\}$  при сдвигах вдоль своих траекторий после десяти шагов по этой диаграмме возвращаются на эту грань  $F_0^4$ .

Для такого отображения Пуанкаре  $\Psi_5: F_0^4 \rightarrow F_0^4$  справедлив аналог леммы 1, из которого следует, что грань  $F_0^4$  содержит в точности две неподвижные точки этого отображения, точку  $E_5 = (1; 1; 1; 1; 1)$  и точку  $\Pi_*^5$ , лежащую во внутренности этой грани. Область  $W_1^5$  содержит в точности один цикл. Обозначим его через  $C_5$ . Этот цикл устойчив и проходит через точку  $\Pi_*^5$  (Аюпова, Голубятников, 2021).

Так же, как и выше, устанавливается аналог леммы 2: матрицы Якоби  $J_4(E_5)$ ,  $J_4(\Pi_*^5)$  и их определители положительны,  $\det J_4(E_5) = 1$ .

Модули собственных чисел матрицы  $J_4(\Pi_*^5)$  не превосходят единицы. В случае, когда эти матрицы Якоби не имеют собственных чисел, по модулю равных единице, построение инвариантной поверхности  $\Sigma^2 \subset W_1^5$  с краем  $C_5$  и инвариантного слоения в области  $W_1^5$  проводится точно так же, как и выше.

В работе (Голубятников, Градов, 2020) установлены условия, при которых неинвариантный страт  $Q^5 \setminus (W_1^5 \cup \{00000\} \cup \{11111\})$  фазового портрета пятимерной системы (7) содержит еще один ее цикл.

Аналогичные построения можно провести и для изучавшегося в работах (Минушкина, 2021; Голубятников, Минушкина, 2022) блочно-линейного аналога шестимерной динамической системы Еловица–Лейблера (Elowitz, Leibler, 2000)

$$\begin{aligned} \dot{m}_1 = L_1(p_3) - k_1 m_1; \dot{p}_1 = \Gamma_1(m_1) - l_1 p_1; \dot{m}_2 = L_2(p_1) - k_2 m_2; \\ \dot{p}_2 = \Gamma_2(m_2) - l_2 p_2; \dot{m}_3 = L_3(p_2) - k_3 m_3; \dot{p}_3 = \Gamma_3(m_3) - l_3 p_3. \quad (8) \end{aligned}$$

Здесь переменные  $m_j$  и  $p_j$  обозначают концентрации трех мРНК и соответствующих им белков TetR, LacI и  $\lambda$ cl (Elowitz, Leibler, 2000; Колесов и др., 2016).

Инвариантная область  $Q^6 = \prod_{j=1}^3 [0, a_j] \times [0, b_j]$ , где  $b_j$  – максимальные высоты ступенчатых функций  $\Gamma_j$ , по-

деленные на коэффициенты  $l_j$ ,  $j = 1, 2, 3$ , разбивается шестью плоскостями  $m_j = 1$ ,  $p_j = 1$ ,  $j = 1, 2, 3$  на 64 блока, из которых конструируется стратификация  $Q^6$  на три подобласти,  $W_1^6$ ,  $W_3^6$ ,  $W_5^6$  с различным качественным поведением траекторий.

Область  $W_5^6$  состоит из 12 блоков, из которых траектории могут перейти в 5 соседних блоков. В симметричном случае, когда  $k_j = l_j = 1$ , в данной подобласти циклов не существует. Однако в  $W_5^6$  содержится двумерная инвариантная поверхность, содержащая кусочно-линейные траектории, которые притягиваются по спирали к точке  $E_6 = (1; 1; 1; 1; 1; 1)$ .

В области  $W_1^6$ , образованной 12 блоками, из которых траектории могут выходить только в один соседний блок, отображение Пуанкаре содержит единственную нетривиальную неподвижную точку  $\Pi_*^6$ , траектория которой является устойчивым предельным циклом для всех траекторий в этой области (Голубятников, Минушкина, 2022).

В области  $W_3^6$ , состоящей из 40 блоков, граф переходов траекторий из блока в блок имеет более сложную комбинаторную структуру. На сегодняшний день переходы траекторий из блока в блок по этой подобласти еще не изучены полностью.

Для гладких аналогов динамической системы (8) установлена единственность стационарной точки (Аюпова и др., 2017). Так же, как и в случае блочно-линейных систем, плоскости, проходящие через стационарную точку и параллельные координатным, разбивают инвариантную область  $Q^6$  на 64 блока. Если матрица линеаризации такой гладкой системы в ее стационарной точке имеет вещественными частями и не имеет мнимых собственных чисел, то в инвариантной области  $W_1^6$  содержится цикл  $C_6$  этой системы (Аюпова и др., 2017). В работе (Кириллова, 2020) получены условия существования в области  $W_1^6$  инвариантной поверхности, ограниченной циклом  $C_6$ .

### Результаты вычислительных экспериментов

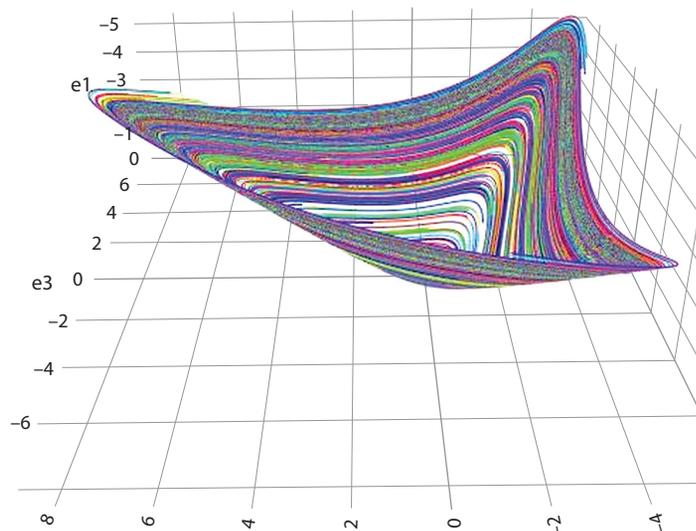
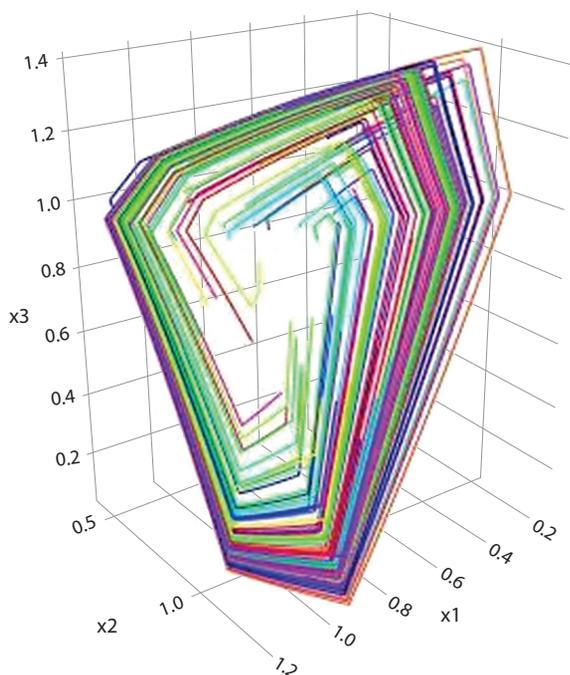
На рисунке слева представлены 100 траекторий динамической системы (3); каждая из них лежит на соответствующем слое слоения в  $W_1^3$ , в окрестности инвариантной поверхности  $\Sigma^2$ . Параметры этой системы имеют значения:  $k_1 = 0.4$ ;  $k_2 = 0.3$ ;  $k_3 = 0.6$ ;  $a_1 = 1.3$ ;  $a_2 = 1.4$ ;  $a_3 = 1.7$ . Начальные данные выбираются случайным образом в прямоугольной окрестности точки  $E_3$ . Справа показаны результаты подобных экспериментов с гладким аналогом системы (3):

$$\frac{dx}{dt} = \frac{10}{1+z^3} - x; \frac{dy}{dt} = \frac{10}{1+x^3} - y; \frac{dz}{dt} = \frac{10}{1+y^3} - z.$$

Здесь хорошо видна ее инвариантная поверхность.

Как показано в статьях (Голубятников и др., 2018; Аюпова, Голубятников, 2021; Golubyatnikov, Minushkina, 2021; Минушкина, 2021), траектории блочно-линейных динамических систем (3), (5), (7), (8) являются кусочно-гладкими, разрывы происходят на плоскостях  $x_j = 1$ , что хорошо видно на рисунке слева.

Для численных экспериментов с траекториями системы (3) разработан программный комплекс на основе язы-



Результаты численных экспериментов с траекториями 3D систем.

ка программирования R (<https://www.r-project.org/>) и пакета Shiny (<https://shiny.rstudio.com/>). Его исходный код доступен на веб-сервисе для хостинга кода GitHub: <https://github.com/AndreyAkinshin/pwLLL>.

Все вычисления выполняются на облачном сервисе. Презентация результатов доступна в веб-браузере по адресу: <https://aakinshin.net/posts/dscs2/>. Используются библиотека ggplot (<https://ggplot2.tidyverse.org/>) и пакет deSolve (<http://desolve.r-forge.r-project.org/>), который содержит методы интегрирования, ранее применяемые нами при численном моделировании других генных сетей. Разработанный интерфейс позволяет задавать все параметры системы (3).

### Заключение

Описано построение семейств инвариантных двумерных поверхностей (слоений) в фазовых портретах блочно-линейных моделей кольцевых генных сетей малых размерностей. Показано, что на каждом слое таких слоений траектории всех точек отталкиваются от границы центральной части фазового портрета и притягиваются к предельному циклу, описывающему осциллирующий режим функционирования соответствующей генной сети. Проведены численные эксперименты, иллюстрирующие доказанную теорему 1.

Для рассматриваемых динамических систем кинетического типа слои инвариантных слоений в фазовых портретах играют роль поверхностей уровней наборов интегралов движения, изучаемых в классической механике (Poincaré, 1892; Арнольд, 2003). Понижение размерностей инвариантных подмножеств фазовых портретов позволяет более наглядно описывать поведение траекторий, в частности значительно упрощает вычислительные эксперименты с такими моделями генных сетей (Likhoshvai et al., 2020). Построение указанных слоений и исследование их

геометрических свойств могут оказаться полезными при изучении динамических характеристик более сложных моделей генных сетей, когда описание свойств большой системы проводится на основе известных результатов о ее подсистемах, имеющих более простое строение.

### Список литературы / References

- Акиншин А.А., Голубятников В.П., Голубятников И.В. О некоторых многомерных моделях функционирования генных сетей. *Сиб. журн. индустр. математики*. 2013;16(1):3-9.
- [Akinshin A.A., Golubyatnikov V.P., Golubyatnikov I.V. On some multidimensional models of gene network functioning. *J. Appl. Ind. Math.* 2013;7(3):296-301. DOI 10.1134/S1990478913030022.]
- Арнольд В.И. Математические методы классической механики. 5-е изд. М., 2003.
- [Arnold V.I. *Mathematical Methods of Classical Mechanics*. 2-d ed. Springer, 1989.]
- Аюпова Н.Б., Голубятников В.П. О единственности цикла в несимметричной трехмерной модели молекулярного репрессилатора. *Сиб. журн. индустр. математики*. 2014;17(1):3-7.
- [Ayupova N.B., Golubyatnikov V.P. On the uniqueness of a cycle in an asymmetric three-dimensional model of a molecular repressor. *J. Appl. Ind. Math.* 2014;8(2):153-157. DOI 10.1134/S199047891402001X.]
- Аюпова Н.Б., Голубятников В.П. Структура фазового портрета одной кусочно-линейной динамической системы. *Сиб. журн. индустр. математики*. 2019;22(4):19-25. DOI 10.33048/sibjim.2019.22.402.
- [Ayupova N.B., Golubyatnikov V.P. Structure of phase portrait of a piecewise-linear dynamical system. *J. Appl. Ind. Math.* 2019;13(4):606-611. DOI 10.1134/S1990478919040033.]
- Аюпова Н.Б., Голубятников В.П. Об одном цикле в пятимерной модели кольцевой генной сети. *Сиб. журн. индустр. математики*. 2021;24(3):19-29. DOI 10.33048/SIBJIM.2021.24.302.
- [Ayupova N.B., Golubyatnikov V.P. On a cycle in a 5-dimensional circular gene network model. *J. Appl. Ind. Math.* 2021;15(3):376-383. DOI 10.1134/S1990478921030029.]
- Аюпова Н.Б., Голубятников В.П., Казанцев М.В. О существовании цикла в одной несимметричной модели молекулярного репрессилатора. *Сиб. журн. вычисл. математики*. 2017; 20(2):121-129. DOI 10.15372/SJNM20170201.

- [Ayupova N.B., Golubyatnikov V.P., Kazantsev M.V. On the existence of a cycle in an asymmetric model of a molecular repressor. *Num. Anal. Appl.* 2017;10(2):101-107. DOI 10.1134/S199542391702001X.]
- Гантмахер Ф.Р. Теория матриц. 3-е изд. М., 1967.  
[Gantmacher F.R. Applications of the Theory of Matrices. New York: London: Interscience Publ., 1959.]
- Глызин С.Д., Колесов А.Ю., Розов Н.Х. Явление буферности в кольцевых генных сетях. *Теор. и матем. физика.* 2016;187(3):560-579. DOI 10.4213/tmf9052.  
[Glyzin S.D., Kolesov A.Yu., Rozov N.Kh. Buffering in cyclic gene networks. *Theor. Math. Phys.* 2016;187(3):935-951. DOI 10.1134/S0040577916060106.]
- Голубятников В.П., Градов В.С. О неединственности циклов в некоторых кусочно-линейных моделях кольцевых генных сетей. *Матем. труды.* 2020;23(1):107-122. DOI 10.33048/mattrudy.2020.23.104.  
[Golubyatnikov V.P., Gradov V.S. Non-uniqueness of cycles in piecewise-linear models of circular gene networks. *Sib. Adv. Math.* 2021;31(1):1-12. DOI 10.3103/S1055134421010016.]
- Голубятников В.П., Иванов В.В. Единственность и устойчивость цикла в трехмерных блочно-линейных моделях кольцевых генных сетей. *Сиб. журн. чист. и прикл. математики.* 2018;18(4):19-28. DOI 10.33048/PAM.2018.18.402.  
[Golubyatnikov V.P., Ivanov V.V. Uniqueness and stability of a cycle in three-dimensional block-linear circular gene network models. *Sibirskii Zhurnal Chistoi i Prikladnoi Matematiki = Siberian Journal of Pure and Applied Mathematics.* 2018;18(4):19-28. DOI 10.33048/PAM.2018.18.402. (in Russian)]
- Голубятников В.П., Иванов В.В., Минушкина Л.С. О существовании цикла в одной несимметричной модели кольцевой генной сети. *Сиб. журн. чист. и прикл. математики.* 2018;18(3):27-35. DOI 10.17377/PAM.2018.18.4.  
[Golubyatnikov V.P., Ivanov V.V., Minushkina L.S. On existence of a cycle in one asymmetric gene network model. *Sibirskii Zhurnal Chistoi i Prikladnoi Matematiki = Siberian Journal of Pure and Applied Mathematics.* 2018;18(3):27-35. DOI 10.17377/PAM.2018.18.4. (in Russian)]
- Голубятников В.П., Минушкина Л.С. Монотонность отображения Пуанкаре в некоторых моделях кольцевых генных сетей. *Сиб. журн. индустр. математики.* 2019;22(3):39-47. DOI 10.33048/sibjim.2019.22.304.  
[Golubyatnikov V.P., Minushkina L.S. Monotonicity of the Poincaré mapping in some models of circular gene networks. *J. Appl. Ind. Math.* 2019;13(3):472-479. DOI 10.1134/S1990478919030086.]
- Голубятников В.П., Минушкина Л.С. О единственности цикла в одной модели кольцевой генной сети. *Сиб. матем. журнал.* 2022;63(1):95-103. DOI 10.33048/smzh.2022.63.106.  
[Golubyatnikov V.P., Minushkina L.S. On uniqueness of a cycle in one circular gene network model. *Sib. Math. J.* 2022;63(1):79-86. DOI 10.1134/S0037446622010062.]
- Иванов В.В. Притягивающий предельный цикл нечетномерной кольцевой генной сети. *Сиб. журн. индустр. математики.* 2022;25(3):25-32. DOI 10.33048/SIBJIM.2022.25.303.  
[Ivanov V.V. Attracting limit cycle of an odd-dimensional circular gene network model. *J. Appl. Ind. Math.* 2022;16(3):409-415. DOI 10.1134/S199047892203005X.]
- Казанцев М.В. О некоторых свойствах графов доменов динамических систем. *Сиб. журн. индустр. математики.* 2015;18(4):42-48. DOI 10.17377/sibjim.2015.18.405.  
[Kazantsev M.V. On some properties of the domain graphs of dynamical systems, *Sibirskii Zhurnal Industrialnoi Matematiki = Siberian Journal of Applied and Industrial Mathematics.* 2015;18(4):42-48. DOI 10.17377/sibjim.2015.18.405. (in Russian)]
- Кириллова Н.Е. Об инвариантных поверхностях в моделях генных сетей. *Сиб. журн. индустр. математики.* 2020;23(4):69-76. DOI 10.33048/SIBJIM.2020.23.405.  
[Kirillova N.E. On invariant surfaces in gene network models. *J. Appl. Ind. Math.* 2020;14(4):666-671. DOI 10.1134/S1990478920040055.]
- Кириллова Н.Е., Минушкина Л.С. О дискретизации фазовых портретов динамических систем. *Изв. АлтГУ.* 2019;108(4):82-85. DOI 10.14258/izvasu(2019)4-12.  
[Kirillova N.E., Minushkina L.S. On the discretization of phase portraits of dynamical systems. *Izvestiya Altayskogo Gosudarstvennogo Universiteta = Izvestiya of Altai State University.* 2019;108(4):82-85. DOI 10.14258/izvasu(2019)4-12. (in Russian)]
- Колесов А.Ю., Розов Н.Х., Садовничий В.А. Периодические решения типа бегущих волн в кольцевых генных сетях. *Изв. РАН. Сер. математическая.* 2016;80(3):67-94. DOI 10.4213/im8398.  
[Kolesov A.Yu., Rozov N.Kh., Sadovnichii V.A. Periodic solutions of travelling-wave type in circular gene networks. *Izvestiya: Mathematics.* 2016;80(3):523-548.]
- Минушкина Л.С. Фазовые портреты блочно-линейной динамической системы в одной модели кольцевой генной сети. *Матем. заметки СВФУ.* 2021;28(2):34-46. DOI 10.25587/SVFU.2021.60.20.003.  
[Minushkina L.S. Phase portraits of a block-linear dynamical system in a model for a circular gene network. *Matematicheskiye Zametki SVFU = Mathematical Notes of NEFU.* 2021;28(2):34-46. DOI 10.25587/SVFU.2021.60.20.003. (in Russian)]
- Хартман Ф. Обыкновенные дифференциальные уравнения. М., 1970.  
[Hartman Ph. Ordinary Differential Equations. New York: John Wiley, 1964.]
- Elowitz M.B., Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature.* 2000;403:335-338. DOI 10.1038/35002125.
- Gaidou Yu.A., Golubyatnikov V.P. On cycles and other geometric phenomena in phase portraits of some nonlinear dynamical systems. In: Geometry and Applications. *Springer Proc. Math. Stat.* 2014;72:225-233. DOI 10.1007/978-3-319-04675-4\_10.
- Glass L., Pasternack J.S. Stable oscillations in mathematical models of biological control systems. *J. Math. Biol.* 1978;6:207-223.
- Golubyatnikov V.P., Minushkina L.S. Combinatorics and geometry of circular gene networks models. *Pisma v Vavilovskii Zhurnal Genetiki i Seleksii = Letters to Vavilov Journal of Genetics and Breeding.* 2020;6(4):188-192. DOI 10.18699/Letters2020-6-24.
- Golubyatnikov V.P., Minushkina L.S. On uniqueness and stability of a cycle on one gene network. *Siberian Electronic Mathematical Reports.* 2021;18(1):464-473. DOI 10.33048/semi.2021.18.032.
- Hastings S., Tyson J., Webster D. Existence of periodic solutions for negative feedback cellular control systems. *J. Diff. Eqn.* 1977;25:39-64.
- Likhoshvai V.A., Golubyatnikov V.P., Khlebodarova T.M. Limit cycles in models of circular gene networks regulated by negative feedback loops. *BMC Bioinformatics.* 2020;21(Suppl. 11):255. DOI 10.1186/s12859-020-03598-z.
- Mallet-Paret J., Smith H. The Poincaré–Bendixson theorem for monotone cyclic feedback systems. *J. Dynam. Diff. Eqns.* 1990;2(4):367-421.
- Poincaré H. Les Méthodes Nouvelles de la Mécanique Céleste. T. I. *Solutions Périodiques. Non-existence des Intégrales Uniformes. Solutions Asymptotiques.* Paris: Gauthier-Villars et fils, 1892.

#### ORCID ID

V.P. Golubyatnikov orcid.org/0000-0002-9758-3833

**Благодарности.** Работа выполнена в рамках государственного задания ИМ СО РАН, проект № FWNF-2022-0009 «Обратные задачи естествознания и задачи томографии».

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 11.08.2022. После доработки 20.09.2022. Принята к публикации 21.09.2022.

Перевод на английский язык <https://vavilov.elpub.ru/jour>

## Апробация технологии оценки мимики лиц для изучения динамики функциональных состояний человека в ЭЭГ-парадигме покоя

А.Н. Савостьянов<sup>1, 2, 3</sup>✉, Е.Г. Вергунов<sup>2</sup>, А.Е. Сапрыгин<sup>1, 2</sup>, Д.А. Лебедкин<sup>2, 3</sup>

<sup>1</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>2</sup> Научно-исследовательский институт нейронаук и медицины, Новосибирск, Россия

<sup>3</sup> Новосибирский национальный исследовательский государственный университет, Гуманитарный институт, Новосибирск, Россия

✉ a-sav@mail.ru

**Аннотация.** В статье представлены результаты исследования, направленного на поиск ковариат для учета деятельности имплицитных когнитивных процессов в условиях функционального покоя испытуемых и при демонстрации им собственного или чужого лица в совместном анализе данных ЭЭГ-эксперимента. Предлагаемый подход основан на анализе динамики мышц лица испытуемого по видео. В пилотном исследовании приняли участие 18 здоровых добровольцев. В эксперименте испытуемые, сидя перед экраном, последовательно закрывали глаза (три пробы по 2 минуты) и открывали их (также три пробы между периодами закрытых глаз) либо перед пустым экраном, либо перед экраном с демонстрацией видеозаписи их собственного лица или лица незнакомого им человека такого же пола, что и участник. У всех испытуемых регистрировали ЭЭГ, ЭКГ и вели запись видео лица. В работе решали отдельную подзадачу эксперимента: апробацию методики оценки динамики активности мышц лица испытуемых по их видео с открытыми глазами для получения ковариат, которые можно включать в последующую обработку совместно с ЭЭГ-коррелятами в нейрокогнитивных экспериментах с парадигмой, не предполагающей выполнение активных когнитивных заданий (resting-state conditions). Показано, что пол испытуемого, статус экрана (пустой, собственное/чужое лицо), номер пробы связаны с различиями в мимической активности лица и могут выступать искомыми ковариатами. Сделан вывод, что анализ динамики мимической активности по видео с открытыми глазами может быть дополнительным методом в нейрокогнитивных исследованиях для изучения имплицитных когнитивных процессов, связанных с восприятием изображения себя и другого, в парадигме функционального покоя.

Ключевые слова: нейрокогнитивные исследования; свое и чужое лицо; ЭЭГ-корреляты; ковариаты; имплицитные когнитивные процессы; самовосприятие.

**Для цитирования:** Савостьянов А.Н., Вергунов Е.Г., Сапрыгин А.Е., Лебедкин Д.А. Апробация технологии оценки мимики лиц для изучения динамики функциональных состояний человека в ЭЭГ-парадигме покоя. *Вавиловский журнал генетики и селекции*. 2022;26(8):765-772. DOI 10.18699/VJGB-22-92

## Validation of a face image assessment technology to study the dynamics of human functional states in the EEG resting-state paradigm

A.N. Savostyanov<sup>1, 2, 3</sup>✉, E.G. Vergunov<sup>2</sup>, A.E. Saprygin<sup>1, 2</sup>, D.A. Lebedkin<sup>2, 3</sup>

<sup>1</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>2</sup> Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia

<sup>3</sup> Institute for the Humanities of Novosibirsk State University, Novosibirsk, Russia

✉ a-sav@mail.ru

**Abstract.** The article presents the results of a study aimed at finding covariates to account for the activity of implicit cognitive processes in conditions of functional rest of the subjects and during them being presented their own or someone else's face in a joint analysis of EEG experiment data. The proposed approach is based on the analysis of the dynamics of the facial muscles of the subject recorded on video. The pilot study involved 18 healthy volunteers. In the experiment, the subjects were sitting in front of a computer screen and performed the following task: sequentially closed their eyes (three trials of 2 minutes each) and opened them (three trials of the same duration between periods of closed eyes) when the screen was either empty or when it was showing a video recording of their own face or the face of an unfamiliar person of the same gender as the participant. EEG, ECG and a video of the face were recorded for all subjects. In the work a separate subtask of the study was also addressed: validating a technique for assessing the dynamics of the subjects' facial muscle activity using the recorded videos of the "eyes open" trials to obtain covariates that can be included in subsequent processing along with EEG correlates in neurocognitive experiments with a paradigm that does not involve the performance of active cognitive tasks ("resting-state conditions"). It was shown that

the subject's gender, stimulus type (screen empty or showing own/other face), trial number are accompanied by differences in facial activity and can be used as study-specific covariates. It was concluded that the analysis of the dynamics of facial activity based on video recording of "eyes open" trials can be used as an additional method in neurocognitive research to study implicit cognitive processes associated with the perception of oneself and other, in the functional rest paradigm.

Key words: neurocognitive studies; own and other face; EEG correlates; covariates; implicit cognitive processes; self-perception.

**For citation:** Savostyanov A.N., Vergunov E.G., Saprygin A.E., Lebedkin D.A. Validation of a face image assessment technology to study the dynamics of human functional states in the EEG resting-state paradigm. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2022;26(8):765-772. DOI 10.18699/VJGB-22-92

## Введение

Технологии нейрокогнитивных исследований чаще всего основываются на использовании различных подходов к регистрации мозговой активности участников обследований при помощи таких методик, как ЭЭГ или фМРТ (Bringas-Vega et al., 2022). В последние два десятилетия (Biswal, 2012; Snyder, Raichle, 2012) большой интерес у исследователей вызывают функциональные состояния мозга, наблюдаемые в условиях отсутствия внешней когнитивной или эмоциональной нагрузки, т. е. в экспериментальной парадигме покоя (resting-state conditions).

В серии исследований было показано, что функциональные состояния мозга в условиях покоя отражают индивидуальные особенности испытуемых, включая их пол (Volf et al., 2015), возраст (Privodnova et al., 2020; Engemann et al., 2022), генетические особенности (Proshina et al., 2018), социокультурную принадлежность (Knyazev et al., 2012), климатогеографические условия проживания (Milakhina et al., 2020), психологические личностные особенности (Kabbara et al., 2020) и предрасположенность к аффективным нарушениям (Greicius et al., 2007). Однако проблемой использования техник нейрокартирования остается большая вариативность фоновых показателей мозговой активности у здоровых испытуемых (Li et al., 2022). Можно привести пример сравнительного исследования M. Li с коллегами, выполненного на выборке из более чем 1500 участников в девяти странах, в котором выявлено, что показатели фоновой ЭЭГ здорового человека значительно различаются в зависимости как от свойств испытуемых, так и от условий регистрации, которые не входят в экспериментальную парадигму (Li et al., 2022). При этом формально одинаковые условия записи (закрытые глаза без внешней нагрузки) могут давать разные результаты в зависимости от того, в каком регионе мира и в какой период года проводилась регистрация ЭЭГ.

Один из факторов, существенно меняющий функциональные состояния мозга в условиях покоя, – наличие или отсутствие размышлений человека о самом себе в период регистрации у него мозговой активности. В работе (Knyazev et al., 2012) показано, что размышления о самом себе индуцируют повышенную активность дефолт-системы мозга. При этом функциональная организация дефолт-системы в этих условиях продемонстрировала достоверные межкультурные различия при сравнении испытуемых из Новосибирска и Тайваня.

В случае с фМРТ дополнительным фактором является реагирование человека на саму ситуацию помещения его в сканер. Запись фМРТ проводится, когда человек лежит в тесном боксе в условиях звукового шума и ограничен-

ной подвижности, иногда требуется введение контраста. Очевидно, что часть людей реагирует на такие условия как на стрессор, тогда как другие люди воспринимают эти условия иначе, что обуславливает сильный разброс оценок функционального состояния испытуемых. Отсюда возникает задача: на основе дополнительных методов найти такие корреляты (или ковариаты), которые при последующем анализе совместно с результатами ЭЭГ- или фМРТ-обследований позволят более полно учесть психофизиологическое состояние испытуемого.

В случае экспериментальной парадигмы с использованием стимулов для индицирования требуемого состояния участников оценка такого состояния происходит в процессе анализа поведенческих показателей (например, таких как точность/скорость реагирования на внешние стимулы), но при исследовании состояния покоя это невозможно.

Другой метод состоит в применении психологических опросников, которые испытуемому предлагается заполнить до или после обследования. Показатели опросников используются как переменные для оценки субъективных состояний человека в условиях эксперимента или его личностных свойств. Однако такой метод ограничен искренностью испытуемого при заполнении опросника и его способностью к адекватной самооценке, что может быть особенно выражено в случае нейropsychиатрических заболеваний.

В нашем пилотном исследовании предлагается подход с использованием ковариат, которые можно получить из динамики активности мышц лица по видео и которые связаны с психофизиологическим состоянием участников ЭЭГ-эксперимента. Анализ активности мышц лица в психофизиологии апробирован (Николаева, Вергунов, 2021), но не использовался для совместного анализа с ЭЭГ-данными.

Мы проверяем гипотезу о том, что динамика активности мимики лица и продолжительность фиксации взгляда испытуемых на экране в пробах при фоновой активности с отсутствием явных экспериментальных заданий различаются в зависимости от таких факторов, как пол испытуемого, демонстрация пустого экрана или экрана с видео его собственного лица или лица другого человека этого же пола, порядок следования пробы.

У испытуемых были проведены комплексное психологическое тестирование для оценки их личностных черт и регистрация ЭЭГ и ЭКГ совместно с записью видео лиц. Однако в рамках данной публикации мы не приводим результаты ЭЭГ, ЭКГ и психометрии, оставляя их для анализа совместно с выявленными ковариатами на последующих этапах эксперимента.

## Материалы и методы

**Описание выборки.** В экспериментах участвовало 18 добровольцев (8 мужчин и 10 женщин, средний возраст  $19.5 \pm 1.3$  года) из числа студентов Новосибирского государственного университета. Перед проведением обследования все участники подписали информированное согласие. Кроме того, все испытуемые заполнили опросник на наличие у них психиатрических или неврологических заболеваний, анкету на самочувствие перед обследованием и на употребление алкоголя или психоактивных веществ. Критериями исключения были:

- наличие установленных медицинских диагнозов;
- употребление наркотиков или психотропных медицинских препаратов;
- состояние алкогольного опьянения или сильного психологического стресса;
- нарушение инструкций при обследовании (прикрытие части лица рукой, резкие движения, изменение позы так, что часть лица выходит из поля видимости камеры и т. д.).

**Процедура обследования.** Во время эксперимента испытуемые сидели в кресле в звукоизолированной камере с приглушенным освещением. На голове участника находился ЭЭГ-шлем, на левой руке и обеих ногах прикреплены электроды для регистрации ЭКГ. Испытуемые были оповещены, что в процессе регистрации ЭЭГ и ЭКГ проводится видеосъемка их лица. Протокол эксперимента был утвержден этическим комитетом НИИ нейронаук и медицины в соответствии с этическими нормами Хельсинкской декларации по проведению биомедицинских исследований.

Участники получали инструкцию минимизировать движения руками, ногами и головой. Они должны были по команде, подаваемой компьютером, открывать или закрывать глаза. От них не требовалось специально сосредотачивать свой взгляд на экране, но и не запрещалось это делать. У каждого участника обследование проводилось в трех различных условиях:

- а) фоновая запись с попеременным закрыванием/открыванием глаз (по 3 пробы каждого типа по 2 минуты), при которой на экране компьютера не было никаких изображений;
- б) запись с открыванием и закрыванием глаз, при которой на экране демонстрировалась видеозапись с изображением лица участника, сделанная ранее в условиях (а) (также по 3 пробы каждого типа по 2 минуты);
- в) запись с открыванием и закрыванием глаз, при которой участнику демонстрировали видеозапись лица незнакомого ему человека того же пола, что и сам участник (также по 3 пробы каждого типа по 2 минуты).

Все участники обследовались во всех трех условиях. Первым условием всегда было условие (а), т. е. запись без дополнительной внешней стимуляции, для половины участников вторым условием было (б) (свое лицо), третьим – (в) (чужое лицо), а для другой половины участников наоборот – вторым было условие (в), а третьим – (б).

В интервалах между фоновыми экспериментами участники выполняли активные экспериментальные задания – решение лингвистических тестов на поиск синтаксических ошибок в предложениях между первым и вторым

фоновым обследованием (примерно 25 мин) и выполнение моторных тестов в парадигме стоп-сигнал (примерно 12 мин) между вторым и третьим фоновым обследованием.

Перед первым экспериментальным условием все участники заполняли русскую версию опросника Ч. Спилбергера для оценки уровня ситуативной тревожности (Ханин, 1976). После завершения первого условия опросник Ч. Спилбергера заполнялся повторно, чтобы оценить, влияет ли участие в обследовании на уровень ситуативной тревожности. Кроме того, после завершения каждого из экспериментальных условий участники заполняли опросник Г.Г. Князева на самочувствие во время записи ЭЭГ (Knyazev et al., 2012). Таким образом, каждый участник заполнял опросник Ч. Спилбергера по два раза (до и после первого экспериментального условия), а опросник Г.Г. Князева – три раза (после каждого экспериментального условия).

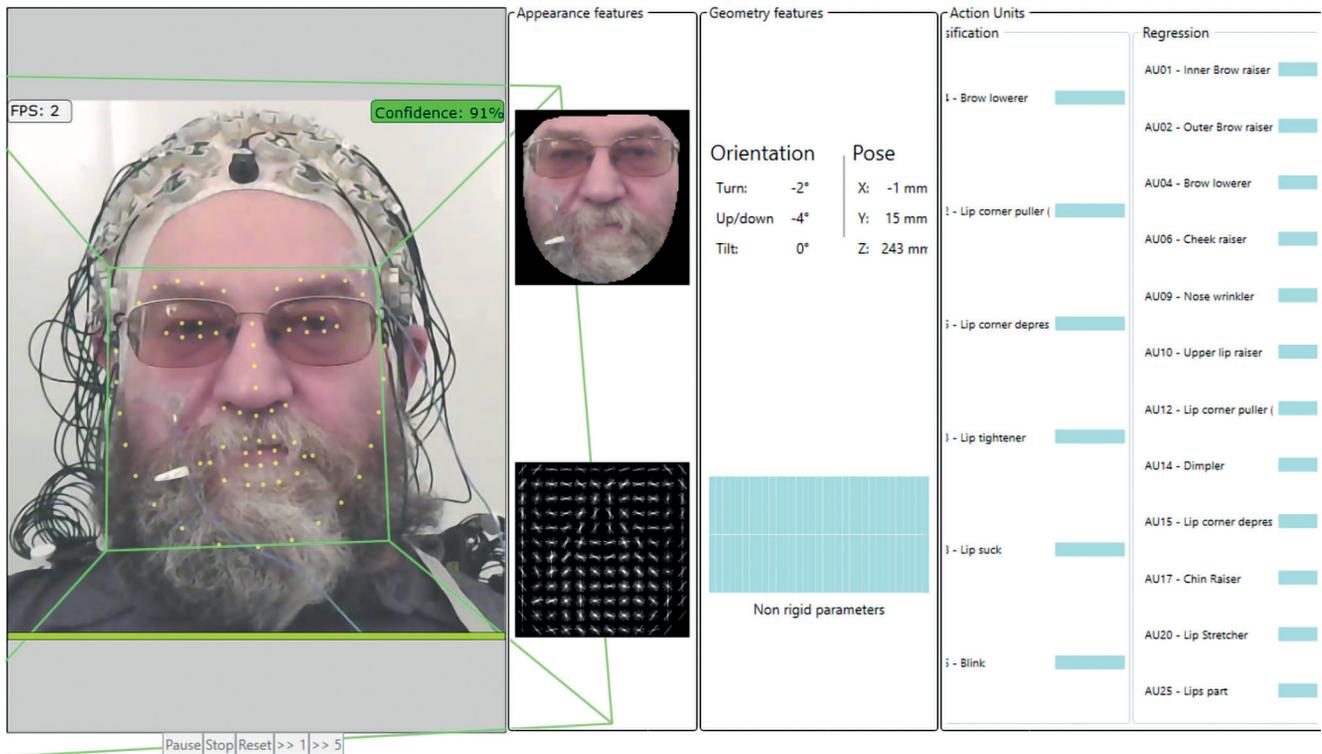
Предлагаемый нами дизайн исследования позволяет контролировать факторы, которые могут сопутствовать имплицитным когнитивным процессам при предъявлении лиц (собственного и чужого) или пустого экрана:

- особенности активности двигательных единиц для мышц лица (AU) согласно facial action coding system (FACS);
- особенности распределения времени в пробе относительно фиксации взгляда испытуемых на экране;
- особенности восприятия испытуемыми разного пола;
- особенности восприятия первой и последующих проб;
- индивидуальная специфика имплицитных когнитивных процессов, связанная с личностными чертами испытуемых, такими как уровень тревожности.

Отметим, что анализ последнего фактора (индивидуальная специфика) не входит в задачи настоящего исследования. Позже совместный анализ результатов психологических опросников с результатами кластеризации по данному фактору будет использован для психофизиологического профилирования испытуемых.

**Методика оценки экспрессии лицевых мышц.** Для оценки экспрессии мышц лица испытуемого по видео широко разрабатываются соответствующие программные инструменты, в том числе общего доступа. В работе был использован фреймворк OpenFace – открытое решение, которое позволяет отслеживать лицо человека по фотографии, по последовательности изображений или по видеопотоку (Сапрыгин et al., 2022). На обычную видеокамеру для компьютера записывали поток видео при выполнении испытуемым задач, затем на основе регрессионной модели выделяли лицевые двигательные единицы (AU) по системе FACS (facial action coding system) и анализировали динамику их активности в период проб с открытыми глазами (рис. 1). Опытным путем было выявлено, что динамике AU в течение периода порядка 2 минут неподвижного сидения испытуемого лучше всего характеризует не значение средней величины или стандартного отклонения (большое количество небольших случайных изменений создает «шум»), а размах значений. Поэтому в анализ был взят именно размах значений экспрессии для каждой AU.

Фреймворк OpenFace основан на использовании подхода CLM (constrained local model). Разрабатываемое ав-



**Рис. 1.** Копия экрана программы с обработкой мимики лица участника обследования из видео с элементами усложнения для анализа (темные очки, борода, усы, шапочка с электродами для регистрации ЭЭГ).

FPS – скорость обработки, число кадров в секунду; confidence – уровень надежности (зеленый цвет говорит о приемлемом уровне); appearance features – распознанные программой особенности мимики после приведения лица в вертикальное положение; geometry features – геометрия 3D-положения и ориентации лица; action units – активность двигательных единиц для мышц лица (AU согласно FACS); classification – значения AU, полученные методом классификации (в настоящем исследовании не используется); regression – значения AU, полученные методом регрессии (см. таблицу); orientation – угловые 3D-показатели ориентации лица; turn – поворот лица (влево+, вправо-); up/down – наклон лица (вверх+, вниз-); tilt – наклон лица к плечу (левому+, правому-); pose – линейные показатели положения лица; X, Y, Z – координаты центра лица (в мм); non rigid parameters – нежесткие параметры; pause, stop, reset, >>1, >>5 – кнопки управления проигрывателем кадров/видео.

торами пилотное программное обеспечение на основе OpenFace\_GUI позволяет визуализировать в реальном времени набор возможностей, которые предоставляют 3D-модели фреймворка OpenFace (координаты ключевых точек лица, положение и углы наклона головы в пространстве, направление взгляда). Фреймворк OpenFace состоит из трех основных частей: 1) программного C++ кода, реализующего основной аналитический поток; 2) файлов предтренированных моделей для детекции лица, детекции и отслеживания ключевых точек лица, вычисления двигательных единиц; 3) Matlab-кода для создания собственных файлов моделей.

Файлы моделей создаются с помощью широкого спектра обучающих наборов данных (датасетов). Программный код фреймворка OpenFace является открытым и доступен по лицензии GNU: <https://github.com/TadasBaltrusaitis/OpenFace>.

**Математические основы модели.** PLS-анализ – это метод получения проекций на латентные структуры (projection to latent structure), первоначальное название которого – «метод частичных наименьших квадратов» (partial least squares). Эффективные инструменты PLS-анализа – бикомпонентные модели (2B-PLS, two-block PLS) (Rohlf, Corti, 2000). Бикомпонентные модели для изучения имплицитных когнитивных процессов выявляют глубинные независимые (ортогональные) «латентные структуры»

(психофизиологические механизмы) одновременно для двух разных блоков (матрицы B1 и B2) многомерных показателей (Kovaleva et al., 2019).

При построении бикомпонентных моделей происходят центрирование рядов данных, масштабирование и повороты обоих блоков для получения максимальной ковариации между матрицами счетов (B1- и B2-score), которые являются проекциями матриц B1 и B2 на искомые латентные структуры. В этом заключается главное отличие 2B-PLS от PCA (principal component analysis, метод главных компонент), который позволяет строить модели только «однокомпонентного» типа. Например, в один блок можно поместить переменные-признаки (состоят только из «0» и «1», дисперсия минимальная), а в другой – ряды инструментальных данных (дисперсия много больше, чем у признаков).

Полученные в бикомпонентной модели латентные структуры описываются с помощью ортогональных матриц нагрузок (B1- и B2-loadings). Строки в матрицах B1 и B2 – это данные по объектам, столбцы – изучаемые показатели. Таким образом, показатели выступают как исходные оси координат (в том числе и коррелированные между собой) и могут рассматриваться как «явные структуры», каждая из которых обуславливает некоторое (обычно небольшое) количество общей дисперсии. Цель бикомпонентной модели состоит в нахождении системы

Блоки переменных для бикомпонентной модели

Переменная	Блок
AU01 – лобная мышца, подниматель медиального края брови	№ 1
AU02 – лобная мышца, подниматель латерального края брови	№ 1
AU04 – мышца гордецов, опускающий брови	№ 1
AU06 – глазничная часть круговой мышцы глаза, подниматель щеки	№ 1
AU09 – подниматель верхней губы и крыла носа, сморщиватель носа	№ 1
AU10 – нижнеглазничная головка квадратной мышцы верхней губы, подниматель верхней губы	№ 1
AU12 – большая скуловая мышца, подниматель уголка губы	№ 1
AU14 – мышца трубачей, щёчная мышца, ямочка на щеке	№ 1
AU15 – треугольная мышца рта, опускающий угол рта	№ 1
AU17 – подбородочная мышца, подниматель подбородка	№ 1
AU20 – мышца смеха, растягиватель губ	№ 1
AU25 – совокупность расслабления подбородочной мышцы и круговой мышцы рта, губы разведены	№ 1
t – доля времени (от продолжительности предъявления) фиксации взгляда на экране испытуемым при предъявлениях стимулов	№ 1
n1–n3 – признак номера предъявления стимула	№ 2
f – признак принадлежности к женскому полу	№ 2
m – признак принадлежности к мужскому полу	№ 2
fn – признак стимула (экран без лица)	№ 2
tf – признак стимула (экран с чужим лицом такого же пола)	№ 2
wf – признак стимула (экран со своим лицом)	№ 2
s6–s49 – признаки индивидуальной специфики (коды испытуемых)	№ 2

Примечание. Классификация AU соответствует Facial Action Coding System (FACS).

пар осей для обоих блоков сразу, которые выражают максимальный шаблон ковариации (Polunin et al., 2019). При этом матрицы нагрузок – это матрицы перехода от исходных «явных структур» к найденным новым «латентным структурам».

В результате применения бикомпонентной модели мы получаем количество латентных структур (новых осей координат), которое равно минимальному числу переменных из двух блоков исходных данных. Заметим, что соотношения для структур сырых данных в блоках остаются теми же самыми после любого количества (и порядка применения) таких операций, как центрирование, масштабирование, поворот, которые применяются в PLS-или PCA-моделях. Таким образом, полностью сохраняется структура сырых данных, в то время как инструменты метода наименьших квадратов (ordinary least squares, OLS) в ряде случаев могут приводить к деформации исходной структуры.

В итоге построения бикомпонентной PLS-модели собирается вся информация из исходных рядов данных (количество которых может достигать многих сотен) в несколько первых независимых латентных структур. Модель 2B-PLS допускает ситуацию, когда число переменных больше (и даже много больше), чем число объектов, а также взаимную коррелированность исходных данных. Более того, ряды данных могут представлять собой линейные комбинации друг друга (Ränner et al., 1994).

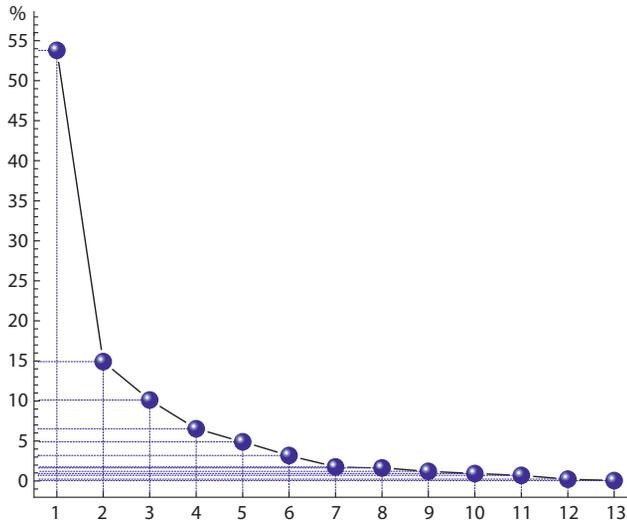
**Результаты и обсуждение**

Построена модель 2B-PLS, в блоки которой вошли переменные, представляющие собой ряды инструментальных данных (13 переменных, блок № 1) и ряды признаков (26 переменных, блок № 2) (см. таблицу). Соответственно, получено 13 латентных структур.

Как следует из рисунка «осыпи» для латентных структур построенной модели 2B-PLS (рис. 2), первый перегиб графика приходится на структуру № 2. Таким образом, структура № 1 (до первого перегиба) будет отражать общие особенности имплицитных когнитивных процессов (обусловленная ею доля наблюдаемой общей дисперсии это подтверждает).

Второй перегиб графика приходится на структуру № 4. Таким образом, для структур № 2 и 3 определяющей будет являться частная специфика имплицитных когнитивных процессов. В последующих структурах растет шумовая компонента – одновременно со снижением доли описываемой общей дисперсии, однако мы еще рассмотрим структуру № 4 – она обуславливает более 5 % общей дисперсии.

Позже анализ результатов психологических опросников совместно с итогами кластеризации по полученным нами структурам может быть использован для целей психофизиологического профилирования испытуемых. Отсюда следует вывод, что для последующей обработки индивидуальных данных испытуемых в ЭЭГ-эксперименте необ-



**Рис. 2.** График осипи для латентных структур построенной модели 2B-PLS.

Ось X – номера латентных структур; ось Y – описываемая ими доля наблюдаемой общей дисперсии.

ходимо оценивать влияние различий в их имплицитных когнитивных процессах, возникающих при появлении на экране своего или чужого лица.

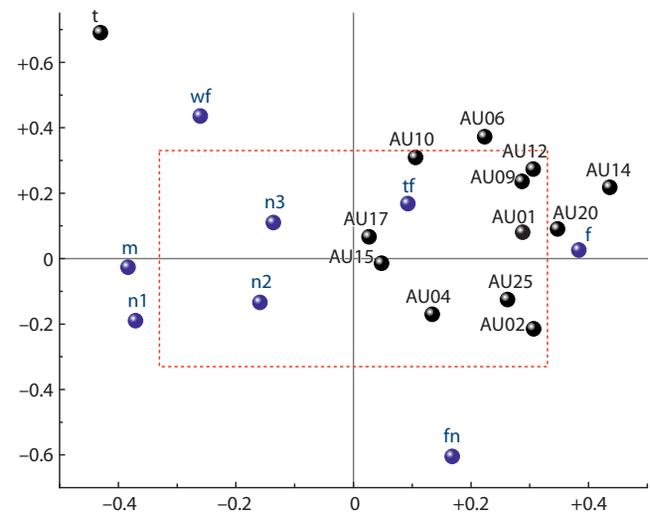
Первые четыре латентные структуры описывают 85.4 % общей дисперсии, и определяющими признаками являются пол, тип стимула и порядок проб.

Согласно рис. 3, первая структура описывает 53.8 % общей дисперсии и определяется долей времени фиксации взгляда испытуемого на экране, активностью «мышцы трубочей» и «мышцы смеха», признаками пола и восприятием всех первых проб. Таким образом, восприятие всех первых проб сопровождается ростом активности мышц «трубочка» и «смеха» и снижением доли времени фиксации взгляда на экране у девушек, а у юношей – наоборот, снижением активности этих мышц и увеличением времени фиксации взгляда на экране.

Вторая структура описывает 14.9 % общей дисперсии и определяется долей времени фиксации взгляда испытуемого на экране, активностью поднимателя щеки и признаками типа стимулов (см. рис. 3). На основании этого можно заключить, что восприятие всеми испытуемыми собственного лица на экране сопровождается ростом активности поднимателя щеки и ростом доли времени фиксации взгляда на экране, а восприятие пустого экрана – наоборот, снижением активности поднимателя щеки и уменьшением времени фиксации взгляда.

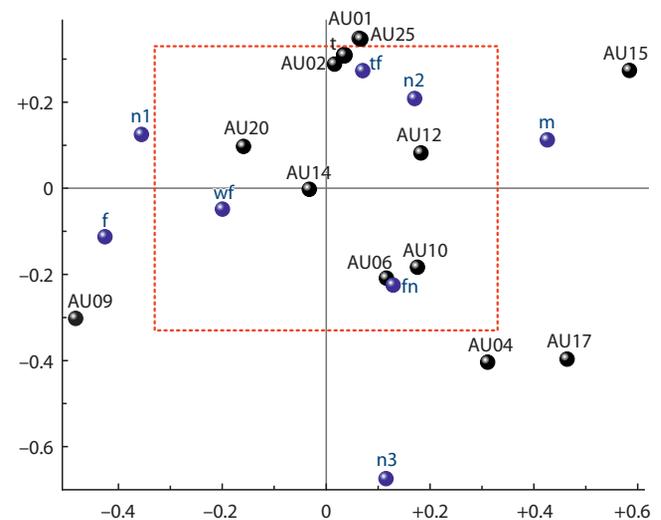
Можно отметить, что в пространстве первых двух латентных структур восприятие чужого лица у всех испытуемых сопровождается ростом активности поднимателей верхней губы и подбородка и опускателя угла рта.

Согласно рис. 4, третья структура описывает 10.1 % общей дисперсии и определяется активностью сморщивателя носа, поднимателя подбородка и признаками пола и первой пробы. Отсюда следует вывод, что восприятие всех первых проб сопровождается ростом активности сморщивателя носа и снижением активности поднимателя подбородка у девушек, а у юношей – наоборот, снижением



**Рис. 3.** Нагрузки (коэффициенты корреляции) переменных для латентной структуры 1 (ось X; 53.8 % общей дисперсии) и структуры 2 (ось Y; 14.9 % общей дисперсии) в модели 2B-PLS.

Здесь и на рис. 4: черный цвет – инструментальные переменные, синий – переменные-признаки (см. таблицы); внутри прямоугольника (красный пунктир) – статистическая достоверность коэффициентов корреляции  $p > 0.05$ ; признаки индивидуальной специфики для улучшения читаемости графика скрыты.



**Рис. 4.** Нагрузки (коэффициенты корреляции) переменных для латентной структуры 3 (ось X; 10.1 % общей дисперсии) и структуры 4 (ось Y; 6.6 % общей дисперсии) в модели 2B-PLS.

активности сморщивателя носа и ростом активности поднимателя подбородка.

Четвертая структура описывает 6.6 % общей дисперсии и определяется признаком последних проб, активностью поднимателя медиального края брови, мышцы гордецов, поднимателя подбородка и разведением губ (см. рис. 4). Из чего можно заключить, что реакция на все третьи пробы у всех испытуемых сопровождается ростом активности поднимателя подбородка и мышцы гордецов, снижением активности поднимателя медиального края брови и степени расслабления подбородочной мышцы и круговой мышцы рта, а также меньшим разведением губ.

Можно отметить, что в пространстве латентных структур № 3 и 4 восприятие чужого лица у всех испытуемых сопровождается ростом доли времени фиксации взгляда на экране и ростом активности поднимателей медиального и латерального края брови, расслаблением подбородочной мышцы и круговой мышцы рта, разведением губ.

Таким образом, в ЭЭГ/ЭКГ-эксперименте в совместную обработку рекомендуется включать (кроме влияния индивидуальных различий имплицитных когнитивных процессов) следующие переменные-ковариаты: пол, порядок следования проб, наличие своего лица на экране/пустой экран.

### Заключение

Электроэнцефалограмма является одним из наиболее распространенных методов для неинвазивного исследования функционального состояния головного мозга человека в норме и при патологиях. При анализе взаимосвязи между показателями ЭЭГ и поведенческой активностью испытуемого в качестве поведенческих метрик обычно выбирают двигательные (существенно реже речевые) ответы испытуемых. Этот выбор обусловлен в первую очередь тем, что такие ответы просто отмечать на записях ЭЭГ. Мы предположили, что изменения в состоянии лицевой мимической мускулатуры могут служить поведенческим фенотипическим признаком, ассоциированным одновременно с личностными особенностями участника обследования, включая его предрасположенность к психическим заболеваниям, и эндофенотипическими параметрами мозговых ритмов.

В данной статье мы предлагаем методологическую идею по регистрации и обработке лицевого видео совместно с регистрацией ЭЭГ. Проведено пилотное исследование по поиску ковариат, статистически связанных с лицевой экспрессией, для их учета в анализе ЭЭГ, регистрируемой в парадигме функционального покоя, а также при демонстрации испытуемым видеоизображения собственного или чужого лица. Этот подход основан на анализе динамики мышц лица испытуемого по видео, которое записывается одновременно с регистрацией ЭЭГ и ЭКГ.

Показано, что динамика активности лицевой мускулатуры отражает контролируемые условия, которые обычно не используются в анализе ЭЭГ-коррелят когнитивных процессов, но, как следует из результатов, могут сопутствовать тем или иным имплицитным когнитивным процессам. Учет таких ковариат, как пол испытуемого, статус экрана (пустой, собственное/чужое лицо) и номер пробы, повысит надежность оценки когнитивного состояния испытуемых и предоставит дополнительную информацию для интерпретации ЭЭГ/ЭКГ-результатов. Кластеризация испытуемых по факторам индивидуальной специфики имплицитных когнитивных процессов составит основу для их эффективного профилирования.

В настоящей работе мы не проводили анализ данных ЭЭГ/ЭКГ и психометрии, поскольку это пилотное исследование с ограниченными задачами. В дальнейшем планируется увеличить размер экспериментальной выборки и провести более детальное сопоставление результатов

анализа активности лицевой мускулатуры с результатами других нейрокогнитивных методик. Для этих перспективных задач мы отработали методику получения данных для профилирования испытуемых по описанным авторами латентным структурам, что позволяет результаты сформированной модели использовать в качестве дополнительных переменных для сводных моделей второго уровня (в том числе результаты ЭЭГ, ЭКГ и др.).

### Список литературы / References

- Николаева Е.И., Вергунов Е.Г. Оценка связи асимметрии лицевой экспрессии с тормозным контролем и латеральными предпочтениями у физически активных мужчин. *Асимметрия*. 2021; 15(4):38-53. DOI 10.25692/ASY.2021.15.4.004.
- [Nikolaeva E.I., Vergunov E.G. Evaluation of the relationship of facial expression asymmetry with inhibitory control and lateral preferences in physically active men. *Asimetriya = Asymmetry*. 2021; 15(4):38-53. DOI 10.25692/ASY.2021.15.4.004. (in Russian)]
- Ханин Ю.Л. Краткое руководство к шкале реактивной и личностной тревожности Ч.Д. Спилберга. Л.: ЛНИИФК, 1976.
- [Khanin Yu.L. Quick Guide to C.D. Spielberger's Scale of State and Trait Anxiety. Leningrad, 1976. (in Russian)]
- Biswal B.B. Resting state fMRI: a personal history. *Neuroimage*. 2012; 62(2):938-944. DOI 10.1016/j.neuroimage.2012.01.090.
- Bringas-Vega M.L., Michel C.M., Saxena S., White T., Valdes-Sosa P.A. Neuroimaging and global health. *Neuroimage*. 2022;260:119458. DOI 10.1016/j.neuroimage.2022.119458.
- Engemann D.A., Mellot A., Hochenberger R., Banville H., Sabbagh D., Gemein L., Ball T., Gramfort A. A reusable benchmark of brain-age prediction from M/EEG resting-state signals. *Neuroimage*. 2022; 262:119521. DOI 10.1016/j.neuroimage.2022.119521.
- Greicius M.D., Flores B.H., Menon V., Glover G.H., Solvason H.B., Kenna H., Reiss A.L., Schlaggar A.F. Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biol. Psychiatry*. 2007;62(5):429-437. DOI 10.1016/j.biopsych.2006.09.020.
- Kabbara A., Paban V., Weill A., Modolo Ju., Hassan M. Brain network dynamics correlate with personality traits. *Brain Connect*. 2020; 10(3):108-120. DOI 10.1089/brain.2019.0723.
- Knyazev G.G., Savostyanov A.N., Volf N.V., Liou M., Bocharov A.V. EEG correlates of spontaneous self-referential thoughts: a cross-cultural study. *Int. J. Psychophysiol*. 2012;86(2):173-181. DOI 10.1016/j.ijpsycho.2012.09.002.
- Kovaleva V.Yu., Pozdnyakov A.A., Litvinov Yu.N., Efimov V.M. Estimation of the congruence between morphogenetic and molecular-genetic modules of gray voles *Microtus s.l.* variability along a climatic gradient. *Ecol. Genet*. 2019;17(2):21-34. DOI 10.17816/ecogen17221-34.
- Li M., Wang Y., Lopez-Naranjo C., Hu S., Reyes R.C.G., Paz-Linares D., Areces-Gonzalez A., Hamid A.I.A., Evans A.C., Savostyanov A.N., Calzada-Reyes A., Villringer A., Tobon-Quintero C.A., Garcia-Agustin D., Yao D., Dong L., Aubert-Vazquez E., Reza F., Razzaq F.A., Omar H., Abdullah J.M., Galler J.R., Ochoa-Gomez J.F., Prichep L.S., Galan-Garcia L., Morales-Chacon L., Valdes-Sosa M.J., Tröndle M., Zulkifly M.F.M., Rahman M.R.B.A., Milakhina N.S., Langer N., Rudych P., Koenig T., Virues-Alba T.A., Lei X., Bringas-Vega M.L., Bosch-Bayard J.F., Valdes-Sosa P.A. Harmonized-Multinational qEEG norms (HarMNqEEG). *Neuroimage*. 2022;256:119190. DOI 10.1016/j.neuroimage.2022.119190.
- Milakhina N.S., Tamozhnikov S.S., Proshina E.A., Karpova A.G., Savostyanov A.N., Afonasiyeva E.B. Delta and gamma activity of resting-state EEG as one of the markers of risk of depressive disorders in migrants of subpolar and polar regions of Siberia. In: 2020 Cognitive Sciences, Genomics and Bioinformatics (CSGB). Novosibirsk, 2020;90-92. DOI 10.1109/CSGB51356.2020.9214596.

- Polunin D., Shtaiger I., Efimov V. JACOBI4 software for multivariate analysis of biological data. *bioRxiv*. 2019;803684. DOI 10.1101/803684.
- Privodnova E.Yu., Slobodskaya H.R., Bocharov A.V., Saprygin A.E., Knyazev G.G. Default mode network connections supporting intra-individual variability in typically developing primary school children: An EEG study. *Neuropsychology*. 2020;34(7):811-823. DOI 10.1037/neu0000699.
- Proshina E.A., Savostyanov A.N., Bocharov A.V., Knyazev G.G. Effect of 5-HTTLPR on current source density, connectivity, and topological properties of resting state EEG networks. *Brain Res*. 2018; 1697:67-75. DOI 10.1016/j.brainres.2018.06.018.
- Ränner S., Lindgren F., Geladi P., Wold S. A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and algorithm. *J. Chemometrics*. 1994;8(2):111-125. DOI 10.1002/cem.1180080204.
- Rohlf F.J., Corti M. Use of two-block partial least-squares to study covariation in shape. *Syst. Biol.* 2000;49(4):740-753. DOI 10.1080/106351500750049806.
- Saprygin A., Lebedkin D., Savostyanov A., Vergounov E. Behavioral and neurophysiological study of subject's personality traits under recognition of sentences about self and others. In: *Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2022)*. Abstracts the Thirteenth International Multiconference, Novosibirsk, 04–08 July 2022. Novosibirsk, 2022;950. DOI 10.18699/SBB-2022-556.
- Snyder A.Z., Raichle M.E. A brief history of the resting state: the Washington University perspective. *Neuroimage*. 2012;62(2):902-910. DOI 10.1016/j.neuroimage.2012.01.044.
- Volf N.V., Belousova L.V., Knyazev G.G., Kulikov A.V. Gender differences in association between serotonin transporter gene polymorphism and resting-state EEG activity. *Neuroscience*. 2015;284:513-521. DOI 10.1016/j.neuroscience.2014.10.030.

---

#### ORCID ID

A.N. Savostyanov [orcid.org/0000-0002-3514-2901](https://orcid.org/0000-0002-3514-2901)  
E.G. Vergunov [orcid.org/0000-0002-8352-5368](https://orcid.org/0000-0002-8352-5368)  
A.E. Saprygin [orcid.org/0000-0001-6789-2953](https://orcid.org/0000-0001-6789-2953)  
D.A. Lebedkin [orcid.org/0000-0002-4356-9067](https://orcid.org/0000-0002-4356-9067)

**Благодарности.** Сбор и обработка данных выполнены при поддержке Российского научного фонда, грант № 22-15-00142 «фМРТ и ЭЭГ корреляты фокуса внимания на собственной персоне как фактора предрасположенности к аффективным расстройствам». Работа А.Н. Савостьянова и А.Е. Сапрыгина по предпроцессингу данных финансировалась из средств бюджетного проекта ИЦиГ СО РАН № FWNR-2022-0020 «Системная биология и биоинформатика: реконструкция, анализ и моделирование структурно-функциональной организации и эволюции генов сетей человека, животных, растений и микроорганизмов».

Авторы благодарят В.Э. Каликина за программную реализацию инструмента для анализа AU по изображению лица на основе фреймворка OpenFace.

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 13.09.2022. После доработки 17.11.2022. Принята к публикации 17.11.2022.

Перевод на английский язык <https://vavilov.elpub.ru/jour>

## Разработка нейронной сети для диагностики риска возникновения депрессии по экспериментальным данным стоп-сигнал парадигмы

М.О. Зеленских<sup>1</sup>, А.Е. Сапрыгин<sup>2, 3</sup>, С.С. Таможников<sup>3</sup>, П.Д. Рудыч<sup>1, 3, 4</sup>, Д.А. Лебедкин<sup>1, 4</sup>, А.Н. Савостьянов<sup>1, 2, 3, 4</sup> 

<sup>1</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

<sup>2</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>3</sup> Научно-исследовательский институт нейронаук и медицины, Новосибирск, Россия

<sup>4</sup> Федеральный исследовательский центр фундаментальной и трансляционной медицины, Новосибирск, Россия

 a-sav@mail.ru

**Аннотация.** В настоящее время возможность спрогнозировать результат развития системы – залог успешного функционирования системы. Повышение качества и объема информации, усложнение ее представления, необходимость обнаруживать скрытые связи делают неэффективным, а чаще всего невозможным, применение классических статистических методов прогнозирования. Среди разнообразных методов прогнозирования особое место занимают методы, основанные на использовании искусственных нейронных сетей. Задачей нашей работы является создание нейронной сети, прогнозирующей риск возникновения депрессии у человека, с помощью данных, полученных при использовании системы тестирования показателей моторного контроля. Стоп-сигнал парадигма (ССП) – это экспериментальный метод, позволяющий оценить способность человека активировать целенаправленные движения или подавлять движения, ставшие неадекватными внешним условиям. В современной медицине СПП чаще всего применяется для диагностики двигательных нарушений, таких как болезнь Паркинсона или последствия инсульта. Мы предположили, что СПП может служить основой для выявления риска развития аффективных заболеваний, включая депрессию. В разрабатываемой нами нейронной сети предполагается комбинирование таких поведенческих показателей, как количество пропущенных ответов, количество правильных ответов, среднее время, количество верных торможений после появления стоп-сигнала. Такой набор показателей обеспечит повышенную точность прогнозирования наличия депрессии у человека. Реализованная в работе искусственная нейронная сеть позволяет по данным, полученным с помощью фиксации реакции на стимулы со стоп-сигналом, диагностировать риск возникновения депрессии. Разработана архитектура и реализована система тестирования показателей моторного контроля у человека, затем протестирована в реальных экспериментах. Проведено сравнение нейросетевых технологий и методов математической статистики. Реализована нейронная сеть для диагностирования риска возникновения депрессии по данным СПП. На примере данных с экспертной оценкой на наличие депрессии и результатов, полученных при использовании системы тестирования показателей моторного контроля, продемонстрирована эффективность нейронной сети (с точки зрения точности).

Ключевые слова: стоп-сигнал парадигма; искусственная нейронная сеть; система тестирования; риск возникновения депрессии; машинное обучение.

**Для цитирования:** Зеленских М.О., Сапрыгин А.Е., Таможников С.С., Рудыч П.Д., Лебедкин Д.А., Савостьянов А.Н. Разработка нейронной сети для диагностики риска возникновения депрессии по экспериментальным данным стоп-сигнал парадигмы. *Вавиловский журнал генетики и селекции*. 2022;26(8):773-779. DOI 10.18699/VJGB-22-93

## Development of a neural network for diagnosing the risk of depression according to the experimental data of the stop signal paradigm

М.О. Zelenskih<sup>1</sup>, А.Е. Saprygin<sup>2, 3</sup>, S.S. Tamozhnikov<sup>3</sup>, P.D. Rudych<sup>1, 3, 4</sup>, D.A. Lebedkin<sup>1, 4</sup>, A.N. Savostyanov<sup>1, 2, 3, 4</sup> 

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>3</sup> Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia

<sup>4</sup> Federal Research Center of Fundamental and Translational Medicine, Novosibirsk, Russia

 a-sav@mail.ru

**Abstract.** These days, the ability to predict the result of the development of the system is the guarantee of the successful functioning of the system. Improving the quality and volume of information, complicating its presentation, the need to detect hidden connections makes it ineffective, and most often impossible, to use classical statistical forecasting methods. Among the various forecasting methods, methods based on the use of artificial neural networks occupy a special place. The

main objective of our work is to create a neural network that predicts the risk of depression in a person using data obtained using a motor control performance testing system. The stop-signal paradigm (SSP) is an experimental technique to assess a person's ability to activate deliberate movements or inhibit movements that have become inadequate to external conditions. In modern medicine, the SSP is most commonly used to diagnose movement disorders such as Parkinson's disease or the effects of stroke. We hypothesized that SSP could serve as a basis for detecting the risk of affective diseases, including depression. The neural network we are developing is supposed to combine such behavioral indicators as: the amount of missed responses, amount of correct responses, average time, the amount of correct inhibition of movements after stop-signal onset. Such a combination of indicators will provide increased accuracy in predicting the presence of depression in a person. The artificial neural network implemented in the work allows diagnosing the risk of depression on the basis of the data obtained in the stop-signal task. An architecture was developed and a system was implemented for testing motor control indicators in humans, then it was tested in real experiments. A comparison of neural network technologies and methods of mathematical statistics was carried out. A neural network was implemented to diagnose the risk of depression using stop-signal paradigm data. The efficiency of the neural network (in terms of accuracy) was demonstrated on data with an expert assessment for the presence of depression and data from the motor control testing system.

Key words: stop signal paradigm; artificial neural network; system for depression risk assessment; machine learning.

**For citation:** Zelenskih M.O., Saprygin A.E., Tamozhnikov S.S., Rudych P.D., Lebedkin D.A., Savostyanov A.N. Development of a neural network for diagnosing the risk of depression according to the experimental data of the stop signal paradigm. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):773-779. DOI 10.18699/VJGB-22-93

## Введение

В настоящее время возможность спрогнозировать результат развития системы является залогом успешного функционирования системы. Повышение качества и объема информации, усложнение ее представления, необходимость обнаруживать скрытые связи делают неэффективным, а чаще всего невозможным, применение классических статистических методов прогнозирования. Особое место занимают методы, базирующиеся на использовании искусственных нейронных сетей.

Основная задача нашей работы – создание нейронной сети, прогнозирующей риск возникновения депрессии у человека, с использованием данных, полученных с помощью системы тестирования показателей моторного контроля (Хайкин, 2006). Все данные взяты из открытой базы ИЦиГ СО РАН (ICBrainDB dataset <https://icbraindb.cytogen.ru/api-v2>).

Группа пациентов с депрессией была обследована на базе клиники НИИ нейронаук и медицины. Наличие большого депрессивного расстройства диагностировалось врачом-психиатром в ходе закрытого интервью на основании критериев МКБ-10. В качестве контрольной группы здоровых людей приглашали участников, которые никогда не проходили лечения в психиатрических клиниках и не обращались к психиатрам за врачебной помощью. Все участники контрольной группы отрицали наличие у них в момент обследования или в течение пяти лет до обследования каких-либо неврологических или психиатрических заболеваний. Кроме того, как пациенты, так и контрольные участники отрицали наличие у них алкогольной или наркотической зависимости и употребление иных психоактивных веществ.

Основные отличия искусственных нейронных сетей от методов математической статистики – параллельная обработка информации и способность к обучению без учителя, т. е. к самообучению (<https://wiki.loginom.ru/articles/normalization.html>). Ниже в виде таблицы представлены результаты сравнения нейронных сетей и методов математической статистики по выбранным критериям (табл. 1).

Устойчивость к лишним данным – важный показатель при работе с большим количеством параметров и отсут-

ствии явных зависимостей, которые мы получаем из данных стоп-сигнал парадигмы (ССП). Самообучение дает возможность выполнять поставленные задачи без стороннего вмешательства, что способствует поиску закономерности между параметрами.

Применение методов математической статистики при поиске зависимостей между стоп-сигнал парадигмой и рисками наличия депрессии не позволяет в полной мере обнаружить их наличие из-за неустойчивости методов к лишним данным, а тем более диагностировать в дальнейшем риск возникновения депрессии у человека. Необходимо использовать технологии нейронных сетей, которые, в отличие от статистических методов, обладают устойчивостью к лишним данным и возможностью для самообучения.

Нейронная сеть должна принимать на вход датасет, состоящий из данных, полученных с помощью ССП, и выводить результат диагностики на риск возникновения депрессии.

Стоп-сигнал парадигма – это экспериментальный метод, позволяющий оценить способность человека активировать целенаправленные движения или подавлять движения, ставшие неадекватными внешним условиям. В современной медицине ССП чаще всего применяется для диагностики двигательных нарушений, таких как болезнь Паркинсона или последствия инсульта. Мы предположили, что ССП может служить основой для выявления риска развития аффективных заболеваний, включая депрессию. В разрабатываемой нами нейронной сети предполагается комбинирование таких поведенческих показателей, как количество пропущенных ответов, количество правильных ответов, среднее время, количество верных стоп-сигналов. Такой набор показателей обеспечит повышенную точность прогнозирования наличия депрессии у человека.

Цель настоящей работы – разработка нейронной сети для диагностики риска возникновения депрессии по данным стоп-сигнал парадигмы. Реализованная в работе искусственная нейронная сеть позволяет по данным, полученным с помощью фиксации реакции на стимулы со стоп-сигналом, диагностировать риск возникновения депрессии.

**Таблица 1.** Сравнение нейронных сетей и математической статистики

Критерий	Нейронные сети	Методы математической статистики
Уровень насыщения	Большой уровень насыщения	Маленький уровень насыщения
Вычислительная мощность	Обладают большой вычислительной мощностью	Обладают меньшей вычислительной мощностью, чем искусственные нейронные сети
Прогрессирование алгоритмов	Постоянное развитие алгоритмов построения искусственных нейронных сетей	Развитие происходит медленно
Отсутствие необоснованного результата	Наличие необоснованных результатов	Отсутствие необоснованных результатов
Затраты времени на разработку	Большое количество времени на разработку	Меньшие затраты времени на разработку
Количество данных для получения результата	Для обучения необходимо большое количество данных	Для использования необходимо меньше данных, чем для искусственных нейронных сетей
Устойчивость к лишним данным	Устойчивы к лишним данным	Неустойчивы к лишним данным
Возможность самообучения	Наличие возможности самообучения	Отсутствие возможности самообучения

**Таблица 2.** Используемый стек технологий

Технология	Мотивация
Язык программирования: Python	На данный момент позволяет проще и быстрее работать с нейронными сетями, чем другие языки программирования (например, Java). Поддерживает широкий спектр библиотек
Библиотека для обработки данных: Pandas – это библиотека с открытым исходным кодом, предоставляющая инструменты для работы с различными структурами данных для языка программирования Python (Виноградова, 2012). Библиотека была использована для парсинга результатов экспериментов и для дальнейшей работы с датасетом	Позволяет обрабатывать необходимые расширения (.dat, .csv, .datx)
Библиотека для построения и работы с графиками: matplotlib – это библиотека для создания таких визуализаций, как гистограммы, столбчатые диаграммы, полосы погрешности, графики когерентности и многое другое (Ivanov et al., 2022). Библиотека была использована для построения графиков потерь во время тренировки и валидации нейронной сети, точности тренировки и валидации	Выбрана из-за возможности построения гистограмм
Библиотека для взаимодействия с искусственными нейронными сетями: Keras – это библиотека глубокого обучения, представляющая собой API, написанное на Python, работающее поверх платформы для машинного обучения TensorFlow ( <a href="https://keras.io/about/">https://keras.io/about/</a> )	Выбрана как одна из наиболее популярных технологий для взаимодействия с нейросетями
Система контроля версий – Github	Выбрана как одна из наиболее популярных систем контроля версий

## Материалы и методы

**Реализация нейронной сети.** В табл. 2 приведены используемые технологии для реализации с аргументацией выбора.

**Архитектура модели.** Для работы с моделью и слоями были использованы классы Sequential и Dense библиотеки Keras.

Класс Sequential представляет собой последовательную архитектуру нейронной сети, что эквивалентно последовательному вызову слоев ([https://keras.io/api/layers/core\\_layers/dense/](https://keras.io/api/layers/core_layers/dense/)).

Класс Dense реализует операцию

$$\text{output} = \text{activation}(\text{dot}(\text{input}, \text{kernel}) + \text{bias}), \quad (1)$$

где activation – поэлементная функция активации, передаваемая в качестве аргумента, kernel – матрица всех весов, созданная слоем, bias – вектор смещения, созданный слоем (<https://keras.io/api/layers/activations/>). Были выделены два слоя:

– слой  $x$ , т.е. слой для работы с объектами, основанными на входных данных, за исключением категории тестируемого;

– слой  $y$ , т.е. слой для работы с ответами, основанными на категории тестируемого.

Для работы со слоем  $x$  была использована функция активации relu. Функция relu – возвращает число, если принимает положительный аргумент, в остальных случаях возвращает 0 (<https://matplotlib.org/stable/index.html>). Для работы со слоем  $y$  была использована функция активации sigmoid, необходимая для вероятностного прогнозирования. Сигмоидная функция активации:

$$\text{sigmoid}(x) = \frac{1}{(1 + \exp(-x))}. \quad (2)$$

Для малых значений функция возвращает значение, близкое к 0, а для больших значений – близкое к 1, причем сигмоида всегда возвращает от 0 до 1 ([https://www.probabilitycourse.com/chapter9/9\\_1\\_5\\_mean\\_squared\\_error\\_MSE.php](https://www.probabilitycourse.com/chapter9/9_1_5_mean_squared_error_MSE.php)).

**Сбор данных для обучения.** В ходе подготовки к разработке нейронной сети был создан сбалансированный датасет на основе данных, полученных при обследовании здоровых людей и пациентов с диагностированной депрессией.

Были выделены следующие входные данные:

- Missed – количество пропущенных ответов у тестируемого;
- Right – суммарное количество правильных ответов у тестируемого;
- Av\_time – среднее время реакции за эксперимент у тестируемого;
- Stop – количество верных игнорирований на стоп-сигнал у тестируемого;
- Practice – количество верных ответов в блоке «Практика» у тестируемого;
- Right\_stop – количество правильных ответов без учета стоп-сигнала;
- Incor\_stop – количество неверных реакций на стоп-сигнал;
- Survive – категория тестируемого (здоров или диагностирована депрессия).

**Подготовка и нормализация данных.** Нормализация данных – это процедура предобработки входных данных, при которой значения признаков, образующих входной вектор, приводятся к заданному диапазону. Нормализация необходима потому, что исходные значения могут изменяться в большом диапазоне и работа нейронной сети с такими данными может привести к некорректному результату (<https://keras.io/api/models/>). Нормализация данных к диапазону [0...1] важна для выставления единой привилегированности признаков, т.е. для выставления одинаковой значимости каждого признака, что позволит сравнивать их между собой в равных условиях.

Для нормализации были выбраны все входные данные датасета, за исключением Survive, так как данный параметр является оценкой и принимает всего два значения, 0 или 1.

**Выбор топологии** искусственной нейронной сети – один из важнейших этапов при использовании нейросетевых технологий для решения практических задач. От этого этапа напрямую зависит адекватность обучения

нейросетевой модели ([https://keras.io/api/models/model\\_training\\_apis/](https://keras.io/api/models/model_training_apis/)). Так как перед нами стоит задача классификации и важно найти любые скрытые связи, то необходимо, чтобы каждый искусственный нейрон был связан с другими нейронами.

Исходя из концепций типов нейронных сетей, был выбран полносвязный тип, поскольку, как упоминалось ранее, каждый искусственный нейрон передает свой выход остальным нейронам.

**Экспериментальный подбор параметров обучения.** В ходе данного этапа разработки нейронной сети необходимо подобрать оптимальные параметры обучения, которые будут демонстрировать наилучшие показатели точности и потерь. Подбор осуществляется путем запуска нейронной сети с возможными параметрами и тестовым датасетом.

В табл. 3 представлены результаты экспериментального подбора параметров обучения, т.е. подбора количества прохождений датасета от начала до конца (epochs) и количества данных на валидацию (validation\_split), на сбалансированном датасете (50 % здоровых, 50 % с диагностированной депрессией, всего 205).

На рис. 1 продемонстрирована точность обучения и валидации при обучении на сбалансированном датасете с выбором epochs = 500 и validation\_split = 0.2.

Таким образом, из-за отсутствия подходящих параметров для дальнейшей работы было решено использовать несбалансированный датасет (65 % здоровых, 35 % с диагностированной депрессией, всего 500).

В табл. 4 приведены результаты экспериментального подбора параметров обучения на несбалансированном датасете.

На рис. 2 показаны точность обучения и валидации при обучении на несбалансированном датасете с epochs = 5000 и validation\_split = 0.2.

Исходя из полученных результатов, было выбрано количество прохождений от начала датасета и до конца (epoch) = 4000, количество данных для валидации (validation\_split) = 0.2.

**Обучение нейронной сети.** Чтобы гарантировать корректность работы искусственной нейронной сети, выборка была поделена на две части: обучающие данные для

**Таблица 3.** Подбор параметров на сбалансированном датасете

Эпохи	Валидация	Точность тренировки	Точность валидации	Потери тренировки	Потери валидации	Вывод
500	0.2	0.1–0.9	0.9–0	0.3–0	0.2–1	Не удовлетворяет
1000	0.2	1	0	0.2–0	0.3–1	Не удовлетворяет
200	0.2	0.1–0.9	0.9–0	0.3–0	0.2–0.8	Не удовлетворяет
100	0.2	1	0	0.13–0.3	0.45–0.8	Не удовлетворяет
1000	0.1	0.2–0.82	1–0	0.31–0	0.2–0.67	Не удовлетворяет
500	0.1	0.18–0.81	1–0	0.3–0	0.2–0.7	Не удовлетворяет
200	0.1	0.82	0	0.1–0	0.45–0.7	Не удовлетворяет
100	0.1	0.19–0.81	1–0	0.35–0	0.1–0.7	Не удовлетворяет
100	0.05	0.2–0.79	1–0	0.27–0.15	0.2–0.45	Не удовлетворяет
100	0.02	0.21–0.78	1–0	0.3–0.2	0.15–0.4	Не удовлетворяет

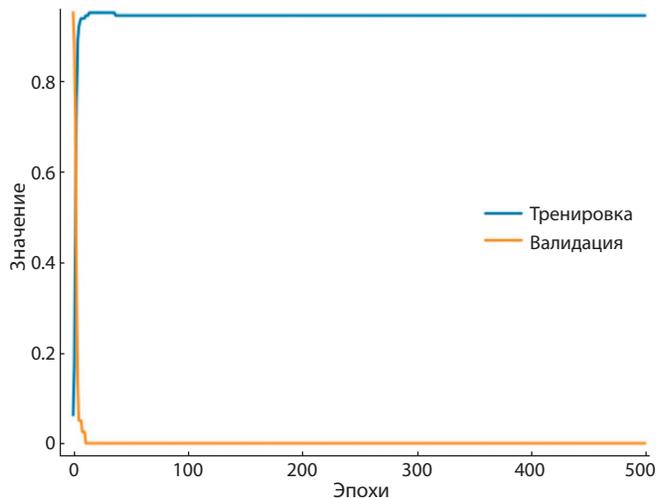


Рис. 1. Пример графика точности обучения и точности валидации при обучении на сбалансированном датасете.

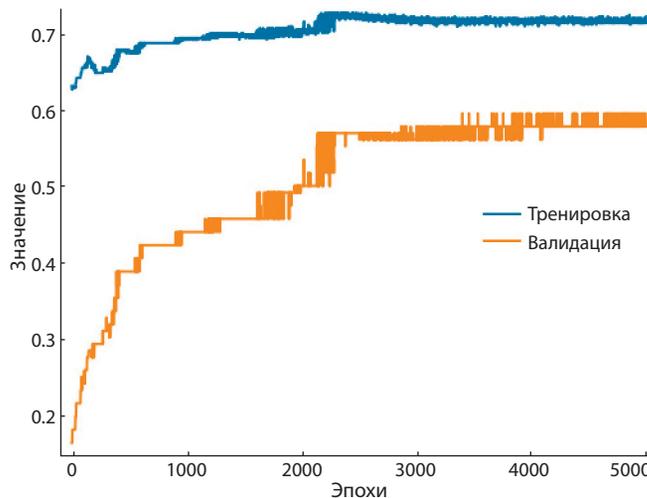


Рис. 2. Пример графика точности обучения и точности валидации при обучении на несбалансированном датасете.

Таблица 4. Подбор параметров на несбалансированном датасете

Эпохи	Валидация	Точность тренировки	Точность валидации	Потери тренировки	Потери валидации	Вывод
1000	0.1	0.6–0.68	0.9–0.85	0.22–0.2	0.16–0.15	Не удовлетворяет
1000	0.05	0.9	0	0	0.9	Не удовлетворяет
1000	0.2	0.13–0.72	0–0.82	0.37–0.2	0.45–0.2	Удовлетворяет
3000	0.2	0.13–0.7	0–0.8	0.37–0.19	0.45–0.2	Удовлетворяет
4000	0.2	0.12–0.73	0–0.84	0.37–0.18	0.45–0.2	Удовлетворяет

Таблица 5. Проверка адекватности обучения

№	1	2	3	4	5	6	7	8
Категория	1	1	0	1	0	0	0	0
Результат	0.767	0.824	0.24	0.927	0.316	0.293	0.276	0.367

Примечание. 0 – с диагностированной депрессией; 1 – без депрессии.

обучения и проверочные данные для проверки работы нейронной сети.

Для обучения были использованы методы `compile` и `fit`. Аргументы метода `compile`: оптимизатор, функция потерь, метрики, веса потерь, список метрик. В методе `fit` аргументы: входные данные, целевые данные, количество выборок, количество эпох, список обратных вызовов, количество данных на валидацию (<https://pandas.pydata.org/pandas-docs/stable/>).

Аргументы, используемые в методе `compile`:

- `loss = "mse"` – среднеквадратическая ошибка:

$$E[(X - \hat{X})^2] = E[(X - g(Y))^2], \quad (3)$$

пусть  $\hat{X} = g(Y)$  – оценка случайной величины  $X$ , учитывающая наблюдение за случайной величиной  $Y$  (<https://www.journaldev.com/45330/relu-function-in-python/>);

- `optimizer = "sgd"` – оптимизатор градиентного спуска с учетом импульса (<https://keras.io/api/optimizers/sgd/>);
- `metrics = ["accuracy"]`.

Аргументы, используемые в методе `fit`:

- $x$  – входные данные;
- $y$  – целевые данные, т. е. оценка;
- `epochs = "4000"` – количество эпох;
- `validation_split = "0.2"` – количество данных на валидацию, используемых в обучающей выборке.

**Проверка адекватности обучения.** Тестирование адекватности обучения проводится на данных, которых не было в обучающих выборках, т. е. используются новые для нейронной сети данные.

Пример последовательности значений (`PSurvived`), полученных в результате работы нейронной сети, с учетом категории данных приведен в табл. 5.

## Результаты

**Технические испытания.** Для технических испытаний нейронной сети были выбраны данные, полученные в ходе экспериментов с использованием нашей системы тестирования показателей моторного контроля у человека

**Таблица 6.** Результаты работы нейронной сети

№	Категория тестируемого	Результат работы нейронной сети	Оценка на основе результата нейронной сети
1	Без экспертной оценки на наличие депрессии	0.8637	Здоровый
2	»	0.5195	Здоровый
3	»	0.6937	Здоровый
4	»	0.7821	Здоровый
5	»	0.7885	Здоровый
6	»	0.4915	Предполагаемо с риском возникновения депрессии
7	»	0.8123	Здоровый
8	»	0.2868	Предполагаемо с риском возникновения депрессии
9	Без экспертной оценки	0.7568	Здоровый
10	С диагностированной депрессией	0.1478	Риск возникновения депрессии – соответствует категории
11	Здоровый	0.9487	Здоровый – соответствует категории
12	С диагностированной депрессией	0.3227	Риск возникновения депрессии – соответствует категории
13	»	0.3114	Риск возникновения депрессии – соответствует категории
14	»	0.2721	Риск возникновения депрессии – соответствует категории
15	»	0.2993	Риск возникновения депрессии – соответствует категории

(без экспертной оценки на наличие депрессии, т. е. без клинического подтверждения), и ранее не использованные данные, которые не участвовали в обучающей выборке (с экспертной оценкой).

Целью технических испытаний является исследование того, как разработанная нейронная сеть будет справляться с классификацией на наличие рисков возникновения депрессии по данным стоп-сигнал парадигмы.

**Входные данные.** Для технических испытаний нейронной сети были выбраны следующие входные данные:

- несбалансированный датасет (0.37 – с диагностированной депрессией, 0.63 – без депрессии);
- максимальное количество пропущенных ответов – 85;
- максимальное суммарное количество правильных ответов у тестируемого – 92;
- максимальное среднее время за эксперимент у тестируемого – 750.0;
- максимальное количество верных игнорирований на стоп-сигнал у тестируемого – 34;
- максимальное количество верных ответов в блоке «практика» у тестируемого – 31;
- максимальное количество правильных ответов без учета стоп-сигнала – 65;
- максимальное количество неверных реакций на стоп-сигнал – 35;
- количество данных для валидации – 0.2, количество эпох – 4000.

**Результаты испытаний.** В табл. 6 описаны результаты работы нейронной сети с оценкой полученных значений.

Таким образом, в ходе технических испытаний получены результаты работы нейронной сети, которые демонстрируют, к какой категории (здоровый/с риском возникновения депрессии) относится тестируемый. По-

лученные показатели полностью соответствуют поставленным диагнозам.

### Заключение

На основе экспериментальных данных, полученных с помощью ССП, был сформирован датасет. Разработана реализация нейронной сети для диагностирования риска возникновения депрессии по данным ССП и далее протестирована. На примере данных с экспертной оценкой на наличие депрессии и данных, полученных с использованием системы тестирования показателей моторного контроля, была продемонстрирована точность классификации нейронной сети. Результаты испытаний в виде показателей качества исполнения нейронной сети описаны ниже:

Показатель	Значение
Потери обучения	0.1657
Точность обучения	0.7821
Потери валидации	0.2415
Точность валидации	0.6667

Методика стоп-сигнал обычно используется для диагностики двигательных нарушений, таких как болезнь Паркинсона, детская гиперактивность или посттравматические расстройства. Ранее для диагностики депрессии методика стоп-сигнала никем не применялась. Мы использовали данную методику в сочетании с методами нейронных сетей и показали, что результаты ССП дают возможность эффективно классифицировать людей на больных депрессией и без депрессии. Необходимо также отметить, что мы не сравнивали больных депрессией с больными с заболеваниями, не связанными с депрессией, неврологическими. Поэтому в данный момент еще не ясно, позволяет ли наш метод разделять больных разными расстройствами на различные подклассы.

## Список литературы / References

- Виноградова Е.Ю. Принципы выбора оптимальной топологии нейронной сети для поддержки принятия управленческих решений. *Управленец*. 2012;7-8:74-78.  
[Vinogradova E.Yu. Principles of choosing the optimal topology of neural network to support managerial decision making. *Upravlenets = The Manager*. 2012;7-8:74-78. (in Russian)]
- Нормализация входных векторов (Normalization) – Loginom Wiki [Электронный ресурс]. URL: <https://wiki.loginom.ru/articles/normalization.html>.  
[Normalization of input vectors (Normalization) – Loginom Wiki [Electronic resource]. URL: <https://wiki.loginom.ru/articles/normalization.html>.]
- Хайкин С. Нейронные сети. Полный курс. М.: Вильямс, 2006.  
[Haykin S. Neural Networks. A Comprehensive Foundation. Moscow: Williams Publ., 2006. (in Russian)]
- About Keras [Электронный ресурс]. URL: <https://keras.io/about/>.  
Dense layer [Электронный ресурс]. URL: [https://keras.io/api/layers/core\\_layers/dense/](https://keras.io/api/layers/core_layers/dense/).  
Layer activation functions [Электронный ресурс]. URL: <https://keras.io/api/layers/activations/>.  
Matplotlib documentation – Matplotlib 3.5.1 documentation [Электронный ресурс]. URL: <https://matplotlib.org/stable/index.html>.  
Mean Squared Error (MSE) [Электронный ресурс]. URL: [https://www.probabilitycourse.com/chapter9/9\\_1\\_5\\_mean\\_squared\\_error\\_MSE.php](https://www.probabilitycourse.com/chapter9/9_1_5_mean_squared_error_MSE.php).  
Model training APIs [Электронный ресурс]. URL: [https://keras.io/api/models/model\\_training\\_apis/](https://keras.io/api/models/model_training_apis/).  
Models API [Электронный ресурс]. URL: <https://keras.io/api/models/>.  
Pandas documentation – pandas 1.4.2 documentation [Электронный ресурс]. URL: <https://pandas.pydata.org/pandas-docs/stable/>.  
ReLU Function in Python – JournalDev [Электронный ресурс]. URL: <https://www.journaldev.com/45330/relu-function-in-python>.  
SGD [Электронный ресурс]. URL: <https://keras.io/api/optimizers/sgd/>.  
Ivanov R., Kazantsev F., Zavarzin E., Klimenko A., Milakhina N., Matushkin Yu., Savostyanov A., Lashin S. ICBrainDB: An integrated database for finding associations between genetic factors and EEG markers of depressive disorders. *J. Pers. Med.* 2022;12(1):53. DOI 10.3390/jpm12010053.

---

### ORCID ID

A.E. Saprygin [orcid.org/0000-0001-6789-2953](https://orcid.org/0000-0001-6789-2953)  
P.D. Rudych [orcid.org/0000-0003-3105-6931](https://orcid.org/0000-0003-3105-6931)  
D.A. Lebedkin [orcid.org/0000-0002-4356-9067](https://orcid.org/0000-0002-4356-9067)  
A.N. Savostyanov [orcid.org/0000-0002-3514-2901](https://orcid.org/0000-0002-3514-2901)

**Благодарности.** Обработка данных при помощи нейронной сети выполнена при финансовой поддержке гранта Российского научного фонда № 22-75-10105. Подготовка базы экспериментальных данных выполнена с участием А.Е. Сапрыгина и А.Н. Савостьянова в рамках бюджетного проекта ИЦиГ СО РАН № FWNR-2022-0020 «Системная биология и биоинформатика: реконструкция, анализ и моделирование структурно-функциональной организации и эволюции генных сетей человека, животных, растений и микроорганизмов».

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 19.09.2022. После доработки 20.11.2022. Принята к публикации 21.11.2022.

Перевод на английский язык <https://vavilov.elpub.ru/jour>

## Программная система на основе 3D симулятора для моделирования эволюции в популяции организмов, обладающих зрительной системой

А.П. Девятериков<sup>1, 2</sup>, А.Ю. Пальянов<sup>1</sup> 

<sup>1</sup> Институт систем информатики им. А.П. Ершова Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>2</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

 palyanov@iis.nsk.su

**Аннотация.** Создание компьютерных моделей, имитирующих работу нервных систем живых организмов с учетом их морфологии и электрофизиологии, – один из важных и перспективных разделов вычислительной нейробиологии. При наличии возможности стремятся моделировать не только нервную систему, но и тело, мышцы, сенсорные системы и виртуальную трехмерную физическую среду, в которой можно наблюдать поведение организма и которая обеспечивает его сенсорные системы адекватными потоками данных, изменяющимися в ответ на движение организма. Для системы из сотен или тысяч нейронов еще можно надеяться задать необходимые параметры и получить функционирование нервной системы, более-менее сходное с таковым для живого организма, как, например, в недавней работе по моделированию головастика *Xenopus*. Однако наибольший интерес, как практический, так и фундаментальный, представляют организмы, обладающие зрением, более сложной нервной системой и, соответственно, значительно более развитыми когнитивными способностями. Определить структуру и параметры нервных систем таких организмов представляется исключительно сложной задачей. Более того, они изменяются с течением времени, в том числе под воздействием воспринимаемых ими потоков сенсорных сигналов и полученного жизненного опыта, включая последствия собственных действий при тех или иных обстоятельствах. Зная структуру нервной системы и число образующих ее нервных клеток хотя бы приблизительно, можно попытаться оптимизировать начальные параметры модели посредством искусственной эволюции, в процессе которой виртуальные организмы будут взаимодействовать и выживать – каждый под управлением собственной версии нервной системы. Помимо этого, эволюционировать могут и правила, по которым мозг изменяется на протяжении жизни организма. Данная работа посвящена созданию нейроэволюционного симулятора, способного осуществлять одновременное функционирование виртуальных организмов, обладающих зрительной системой, которые взаимодействуют между собой. Приведены расчеты, показывающие, сколько вычислительных ресурсов требуется для работы моделей физического тела организма, нервной системы и виртуальной среды обитания, а также определена производительность симулятора на современной настольной вычислительной системе в зависимости от числа одновременно моделируемых организмов.

Ключевые слова: нервная система; зрительная система; виртуальный организм; популяция; компьютерное моделирование; нейроэволюционный симулятор.

**Для цитирования:** Девятериков А.П., Пальянов А.Ю. Программная система на основе 3D симулятора для моделирования эволюции в популяции организмов, обладающих зрительной системой. *Вавиловский журнал генетики и селекции*. 2022;26(8):780-786. DOI 10.18699/VJGB-22-94

## A software system for modeling evolution in a population of organisms with vision, interacting with each other in 3D simulator

A.P. Devyaterikov<sup>1, 2</sup>, A.Yu. Palyanov<sup>1</sup> 

<sup>1</sup> A.P. Ershov Institute of Informatics Systems of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

 palyanov@iis.nsk.su

**Abstract.** Development of computer models imitating the work of the nervous systems of living organisms, taking into account their morphology and electrophysiology, is one of the important and promising branches of computational neurobiology. It is often sought to model not only the nervous system, but also the body, muscles, sensory systems, and a virtual three-dimensional physical environment in which the behavior of an organism can be observed and which provides its sensory systems with adequate data streams that change in response to the movement of the organism. For a system of hundreds or thousands of neurons, one can still hope to determine the necessary parameters and get the functioning of the nervous system more or less similar to that of a living organism – as, for example, in a recent work on the modeling of the *Xenopus* tadpole. However, of greatest interest, both practical and fundamental, are organisms

that have vision, a more complex nervous system, and, accordingly, significantly more advanced cognitive abilities. Determining the structure and parameters of the nervous systems of such organisms is an extremely difficult task. Moreover, at the cellular level they change over time, these including changes under the influence of the streams of sensory signals they perceive and the life experience gained, including the consequences of their own actions under certain circumstances. Knowing the structure of the nervous system and the number of nerve cells forming it, at least approximately, one can try to optimize the initial parameters of the model through artificial evolution, during which virtual organisms will interact and survive, each under the control of its own version of the nervous system. In addition, in principle, the rules by which the brain changes during the life of the organism can also evolve. This work is devoted to the development of a neuroevolutionary simulator capable of performing simultaneous functioning of virtual organisms that have a visual system and are able to interact with each other. The amount of computational resources required for the operation of models of the physical body of an organism, the nervous system and the virtual environment was estimated, and the performance of the simulator on a modern desktop computing system was determined depending on the number of simultaneously simulated organisms.

Key words: nervous system; vision system; virtual organism; population; computational modeling; neuroevolution simulator.

**For citation:** Devyaterikov A.P., Palyanov A.Yu. A software system for modeling evolution in a population of organisms with vision, interacting with each other in 3D simulator. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):780-786. DOI 10.18699/VJGB-22-94

## Введение

Компьютерные модели, основанные на данных об электрофизиологии и морфологии нервных систем живых организмов и имитирующие их работу, являются мощным инструментом в нейробиологии. С их помощью можно на основе знаний и представлений о функционировании отдельных нервных клеток и о механизмах взаимодействия между ними рассчитывать динамику активности сетей нервных клеток. Модель мышечной системы может функционировать в комплексе с моделью тела организма, оснащенного нервной и сенсорной системами, которое помещено в виртуальную трехмерную физическую среду, что предоставляет исследователю значительные преимущества. Во-первых, это позволяет наблюдать и регистрировать как поведение модели тела организма, так и активность нервной системы, вплоть до активности отдельных клеток, их отростков и синапсов. Во-вторых, модель нервной системы получает от виртуальной окружающей среды поток сигналов, изменяющихся в ответ на действия организма, осуществляемые посредством мышечной системы, управляемой его «мозгом», т. е. имеется постоянная обратная связь между действиями и их последствиями, как и в реальности. Одной из целей такого моделирования является проверка адекватности моделей нервных клеток посредством сравнения активности нервной системы и поведения реального организма и его виртуального двойника.

Широкую известность в этой связи получил один из наиболее простых многоклеточных организмов, беспозвоночное *Caenorhabditis elegans*, нервная система которого состоит всего из 302 нейронов (Sarma et al., 2018). Достаточно убедительного сходства оригинала и модели удалось добиться для головастика лягушки *Xenopus*, нервная система которого была представлена сетью, состоящей примерно из 2300 нейронов (Ferrario et al., 2021). Однако ни *C. elegans*, ни головастик *Xenopus* на двухдневной стадии развития не обладают зрительной системой.

Попытки моделирования значительно более сложных организмов, таких как мышь (~70 млн нейронов (Herculano-Houzel et al., 2006)) или крыса (~200 млн нейронов (Herculano-Houzel, Lent, 2005)), и их нервных систем

тоже предпринимались, но создать виртуальные организмы-двойники на их основе пока не удалось. Ведется также работа по «оцифровке» и моделированию нервной системы плодовой мушки дрозофилы (~100 тыс. нейронов (Scheffer et al., 2020)). Еще один исключительно перспективный объект моделирования – муравьи (~250 тыс. нейронов (Moffet et al., 2021)). Эти насекомые обладают неподвижными фасеточными глазами, состоящими из 100–3000 омматидиев (структурные и функциональные единицы таких глаз, количество которых зависит от вида муравья и его специализации), обеспечивающими цветное зрение с довольно скромным разрешением – от 10×10 до 55×55 «пикселей». Так, например, у муравьев вида *Myrmica ruginodis* глаза обычно включают от 109 до 169 омматидиев, у *Camponotus crassus* и *Pseudomyrmex adustus*, ведущих дневной образ жизни, – до 700 и 930 соответственно (Aksoy, Camlitepe, 2018), а максимально известное их число для муравьев, около 3000, зарегистрировано у тропического вида *Gigantiops destructor* (Macquart et al., 2006).

Примечательно, что муравьи являются самыми простыми организмами, успешно проходящими зеркальный тест, т. е. они способны отличить собственное отражение в зеркале от другого муравья, которого они могут видеть сквозь обычное прозрачное, не зеркальное стекло такого же размера (Cammaerts M.-C., Cammaerts R., 2015). Принцип проведения зеркального теста состоит в том, чтобы о нем упомянуть. Перед зеркалом муравьи приводят себя в порядок или совершают необычные движения головой и антеннами, чего не наблюдается, когда они видят за стеклом сородичей. Если муравью на переднюю часть головы нанести небольшую метку, к примеру голубого цвета, то он, увидев себя в зеркале, будет стараться избавиться от нее, пытаться счистить с помощью лапок. Если же метка такого же цвета, как и тело муравья, или если она нанесена на затылочную часть, не видимую в зеркале, то муравей не проявляет беспокойства и попыток счистить ее. Таким образом, муравьи замечают на себе метку и ведут себя так, как будто понимают, что она именно на себе, а не на другом муравье, ориентируясь лишь на зрительные сигналы.

Значительный научный интерес представляет моделирование как одного муравья с меткой или без нее возле зеркала, так и одновременно нескольких муравьев, способных видеть друг друга, взаимодействовать между собой и с окружающими предметами. Ориентация на местности у муравьев также осуществляется в основном посредством зрения (Buehlmann et al., 2020).

Каковы требования для программной системы и вычислительного оборудования, с помощью которых можно осуществлять компьютерное моделирование среды обитания и жизнедеятельности группы муравьев – виртуальных организмов, включающих тело, мышечную, сенсорную и нервную системы? Предполагается, что организмы могут взаимодействовать друг с другом в физическом мире и «видеть» друг друга, т. е. их нервная система получает на вход поток видеоданных, соответствующих зрительно-му восприятию «от первого лица». Задача «оцифровки» структуры нервной системы, включая трехмерную морфологию каждого нейрона, его отростков и синапсов, является исключительно трудоемкой, однако, возможно, этого и не требуется, поскольку мозг все равно достаточно пластичен и в течение жизни организма претерпевает заметные структурные изменения, в том числе и у муравьев (Penick et al., 2021). При этом о закономерностях, лежащих в основе изменений мозга, известно мало. Поэтому имеет смысл ставить задачу о моделировании организма, обладающего телом и сенсорными системами муравья (по меньшей мере зрительной и механосенсорной, а также простейшими обонятельными и вкусовыми рецепторами) и нервной системой с близким числом нейронов и синапсов. Насколько быстро может осуществляться такое моделирование и можно ли рассчитывать на то, что виртуальная эволюция в такой системе поможет искусственным нейронным сетям достичь когнитивных возможностей, при которых виртуальные организмы смогут эффективно выживать, решая для этого более или менее сложные задачи, связанные с поиском пищи, избеганием опасностей и другой деятельностью?

## Материалы и методы

**Программная система.** В соответствии с обозначенной проблематикой, для решения поставленных задач используется компьютерное моделирование, производимое с помощью разработанного нами программного обеспечения для проведения численных экспериментов в области нейроэволюционного моделирования. При разработке моделей был выбран современный 3D движок Unigine (unigine.com), используемый для создания игр, систем виртуальной реальности, программ интерактивной визуализации, обучающих систем в различных сферах деятельности и др., с поддержкой платформ Windows и Linux.

Модуль физической симуляции поддерживает определение столкновений, физику твердого тела, различные типы сочленений (шарнирные, шаровые, призматические, цилиндрические и др.), динамическое разрушение объектов, ткани, плавание объектов, поля сил, обратный ход времени и др. (<https://developer.unigine.com/ru/docs/latest/principles/physics/>). Имеется возможность использовать в симуляции зеркала, что в будущем может пригодиться

для проведения «зеркального теста». В Unigine встроены язык C++, что позволяет использовать при разработке собственный программный код – например, для моделирования сетей нейронов, управляющих виртуальными организмами и получающих сигналы от их сенсорных систем.

**Модель тела «муравья».** Простейшая модель тела «муравья», которую мы спроектировали и использовали в качестве первого приближения для оценки производительности симулятора, представлена на рис. 1. В дальнейшем планируется разработать и использовать значительно более детализированную и реалистичную версию.

В простейшей тестовой сцене на плоскости случайным образом расположены частицы пищи (показаны зеленым цветом) и несколько десятков виртуальных организмов (рис. 2).

**Зрительная система.** На рис. 3 приведены примеры изображений, воспринимаемых «видеокамерой», расположенной на голове организма и направленной вперед (на данный момент реализовано только цветное монозрение, хотя на будущее запланировано и стерео). Был выбран размер изображений  $30 \times 30$  пикселей, приблизительно равный среднему пространственному разрешению среди рассмотренных выше примеров зрительных систем муравьев. Так как сами изображения достаточно малы, для удобства восприятия на рисунке они пропорционально увеличены в 5 раз (одноцветный квадрат размером  $5 \times 5$  пикселей соответствует одному реальному пикселю-«рецептору»).

Изображение задается тремя матрицами, каждая из которых представляет отдельный цветовой канал (красный – R, зеленый – G, синий – B). Каждая матрица имеет размер  $30 \times 30$ . Из них формируется массив данных, *Input*, состоящий из 2700 элементов и организованный следующим образом:

$$\begin{aligned} \text{Input}(r) &= R(i, j), \quad r = i \cdot 30 + j, \\ \text{Input}(g) &= G(i, j), \quad g = i \cdot 30 + j + 900, \\ \text{Input}(b) &= B(i, j), \quad b = i \cdot 30 + j + 1800, \end{aligned}$$

где  $0 \leq i < 30, \quad 0 \leq j < 30$ .

У симуляции есть некоторая частота обновления кадров, зависящая от вычислительных мощностей компьютера, сложности моделируемой сцены и количества «муравьев». С заданной периодичностью у каждой особи формируется такой массив, содержимое которого поступает на вход «нервной системе» организма.

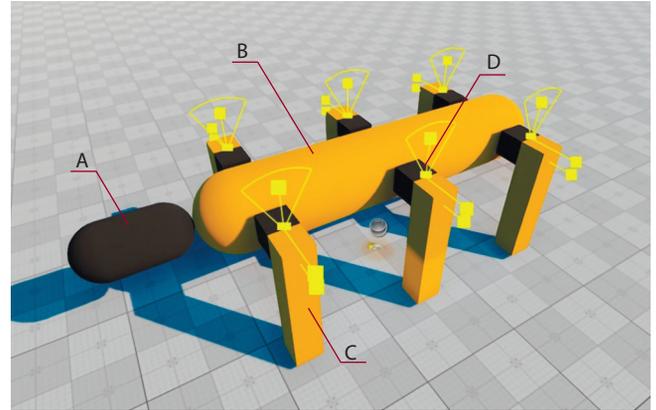
**Нервная система.** Зрительные сигналы поступают в «нервные системы» виртуальных организмов, которые в самом начале симуляции, для первого поколения «муравьев», представляют собой случайно сгенерированные сети из искусственных нейронов, подобных тем, что использовались в перцептронах (Розенблатт, 1965) для распознавания букв, цифр и геометрических фигур. В нашем случае число нейронов в каждой из сетей составляло около 3000. В пределах жизни одной особи сети имеют статическую топологию. Перцептрон состоит из S-элементов (сенсорных), одного или нескольких слоев A-элементов (ассоциативных) и R-элементов (реагирующих). A-элементы задаются набором матриц весов  $A_1, A_2, \dots, A_n$  и векторов сдвига  $b_1, b_2, \dots, b_n$ . На массив

*Input*, описанный выше, матрицы и вектора действуют следующим образом:

$$result_i = A_i \cdot result_{i-1} + b_i,$$

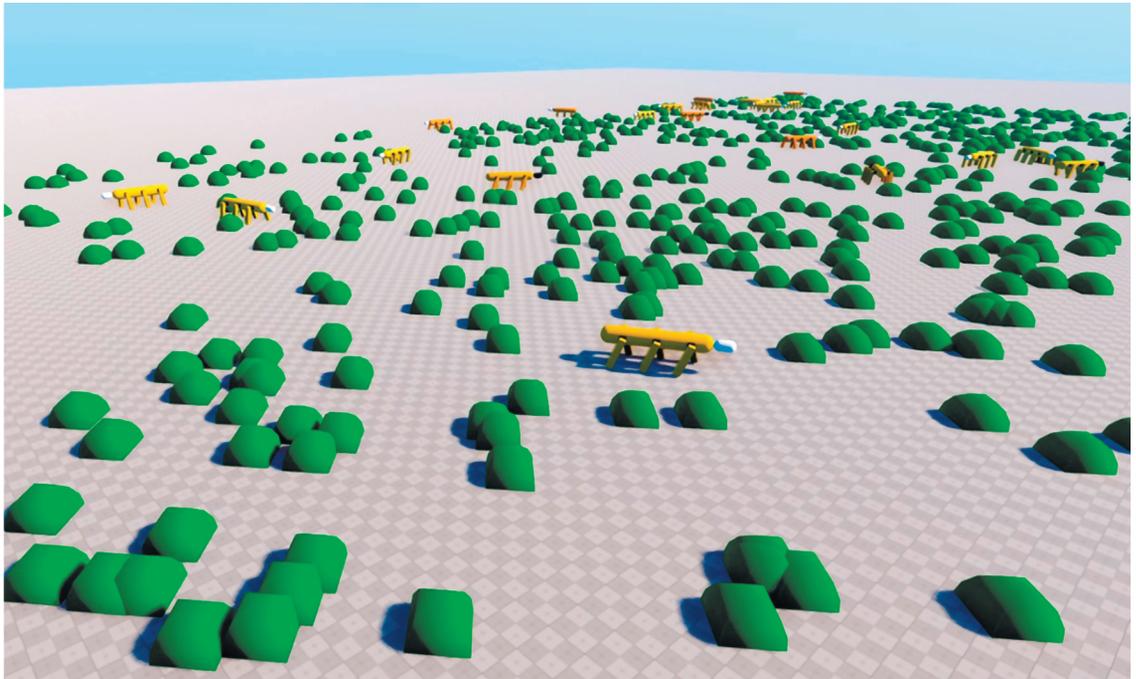
где  $result_0$  – слой из сенсорных элементов, содержащий массив воспринимаемой «муравьем» зрительной информации,  $i = 1, \dots, n$ . Если в результате активируются те или иные R-элементы, это приводит к соответствующим действиям, выполняемым муравьем (изменение скорости, поворот влево или вправо).

**Моделирование эволюции.** Некоторые наборы весов описанных выше перцептронов обеспечивают более эффективное выживание, т.е. способность воспринимать поток зрительных сигналов «от первого лица», анализировать их и управлять движением организма таким образом, чтобы регулярно достигать частиц пищи и поддерживать необходимый «уровень энергии» в организме (сытое состояние). Те организмы, которые слишком долго остаются



**Рис. 1.** Простейшая 3D модель тела «муравья», общий вид.

A – голова, B – туловище, C – ноги, D – подвижные соединения, связывающие ноги и туловище. Голова имеет подвижное соединение с туловищем.



**Рис. 2.** Общий вид симуляции – тестовая сцена с несколькими десятками виртуальных организмов.



**Рис. 3.** Вид «от первого лица», несколько примеров.

На первом примере (слева) можно различить другую особь (вверху, в коричневых тонах) и тень воспринимающего это изображение виртуального организма (темно-серого цвета).

голодными, вымирают, а «долгожители» имеют возможность порождать потомков, наследующих структуру их нейронных сетей. В настоящее время в симуляторе половое размножение не реализовано, потомок порождается лишь одним родителем. В природе такой механизм размножения, называемый партеногенезом, тоже встречается, в том числе у множества типов членистоногих, включая, кстати, 8 видов муравьев, а также примерно у 70 видов позвоночных.

В симуляторе текущий «уровень энергии» организма обозначен как «сытость» –  $Satiety(t)$ , с которой связаны следующие величины:

$MaxSatiety$  – максимальная сытость организма (по умолчанию равна 15);

$BirthSatiety = MaxSatiety \cdot 0.7$  – сытость организма, при достижении которой он порождает потомка. При этом половина имеющихся ресурсов остается у организма, а половина переходит потомку.

Каждый организм инициализируется с сытостью  $Satiety(0) = 8$ . Каждый раз по прошествии заданного интервала времени он теряет единицу сытости (в процессе жизнедеятельности организма «расходуется энергия»). При  $Satiety(t) = 0$  организм умирает. При поглощении пищи организм получает единицу сытости, пока не достигнет  $MaxSatiety$ .

При генерации потомка он наследует нейросеть родителя с изменениями, которые осуществляются по следующим правилам:

$\epsilon, \delta$  – случайные величины с равномерным распределением;

$\epsilon \in [a, b]$  – вероятность мутации нейрона,  $0 \leq a \leq b \leq 1$ ;  
 $\delta \in [c, d]$  – величина изменения веса в элементе матрицы при мутации,  $c \leq d$ . Значения  $a, b, c$  и  $d$  при необходимости могут быть изменены пользователем.

Каждый элемент  $A_k(i, j)$  и  $b_k(l)$  ( $k = 1, \dots, n$ ) матриц весов и векторов сдвига меняется на  $+\delta$  или  $-\delta$  с вероятностью  $\epsilon$ .

## Результаты

Основным результатом на данном этапе работы являются созданный нами прототип симулятора (включающий трехмерный физический мир, модель физического тела муравья, модель зрительной системы и модель нервной системы), исходные программные коды которого доступны в репозитории по ссылке (<https://github.com/NotNa19/AntPrototype>), а также оценка его скорости работы на различных вычислительных системах в зависимости от их характеристик и от числа нейронов в нервной системе виртуальных организмов. Дальнейшее развитие проекта зависит от возможности производить нейроэволюционное моделирование как минимум одного, а желательно нескольких виртуальных организмов, «нервная система» которых сравнима с таковой для настоящего муравья по числу нервных клеток.

Ниже приведены характеристики вычислительных систем, на которых производилось тестирование, и максимальное количество виртуальных организмов, моделируемых одновременно, в пределах которого работа симулятора оставалась стабильной. В данном случае под «стабильной работой» подразумевается корректное функ-

ционирование организмов и их физических тел. Дело в том, что в текущей версии Unigine при низкой частоте кадров могут происходить задержки между движением различных составляющих организма, может не срабатывать обработка коллизий между объектами «организм» и «пища», а также возникать другие проблемы подобного рода. Их решение является возможным и запланировано на будущее, но требует более глубокого изучения механизмов 3D движка. При разрешении экрана  $1920 \times 1080$  пикселей и частоте обновления кадров (frames per second, FPS) не менее 30 за секунду работа симулятора была стабильной, однако число одновременно моделируемых особей оказывало влияние на эти показатели. На имеющемся в нашем распоряжении вычислительном оборудовании были получены следующие значения:

Характеристики вычислительной системы	Максимальное число виртуальных организмов, при котором работа симулятора стабильна
CPU Intel Core i5-7300HQ 2.50 GHz GPU GeForce GTX1050 Ti, 4 Gb	50
CPU AMD Ryzen 7 2700X 3.70 GHz GPU NVIDIA GeForce 1060, 6 Gb	80
CPU AMD Ryzen 5 5600X 3.7/4.6 GHz GPU MSI GeForce RTX 3060 Ti, 8 Gb	150

Детализация времени, затрачиваемого на различные этапы работы симулятора, показала, что при небольшом размере нервной системы (тысячи–десятки тысяч нейронов) наиболее существенным фактором, ограничивающим скорость его работы, является процесс получения зрительной информации «от первого лица» для популяции муравьев, даже с учетом того, что многопоточность расчетов обеспечивается самим движком. Зависимость максимального числа особей в симуляции от количества нейронов в «нервной системе» виртуального организма (у всех особей одинаковое число) тоже была исследована. Для GeForce RTX 3060 Ti + AMD Ryzen 5 5600X получены следующие показатели:

Число нейронов	Максимальное число виртуальных организмов, при котором работа симулятора стабильна
3000	150
10000	50
100000	10

Достаточно ощутимое влияние на производительность системы оказывают также затраты на визуализацию для внешнего наблюдателя. Мы произвели соответствующие измерения и получили следующие данные для конфигурации, включающей CPU AMD Ryzen 7 2700X 3.70 GHz и GPU NVIDIA GeForce 1060 6 Gb:

- при выполнении симуляции с пустой сценой (с визуализацией для внешнего наблюдателя или без нее) получается стабильно 9000 тактов за 60 с (в среднем 150 тактов/с);
- при выполнении симуляции с 80 организмами с визуализацией для внешнего наблюдателя получаем 5400 тактов за 60 с (в среднем 90 тактов/с), а без визуализации – 7800 тактов за 60 с (в среднем 130 тактов/с);
- при большей нагрузке (100 особей и большее количество еды) получилось 1800 тактов за 60 с с визуализа-

цией (в среднем 30 тактов/с) и 4500 тактов за 60 с без визуализации (в среднем 75 тактов/с).

Таким образом, визуализация для внешнего наблюдателя (пользователя) действительно играет достаточно существенную роль в общей производительности системы и ее стоит производить лишь в случаях, когда она действительно необходима: например, при отладке или при записи демонстрационных видеороликов, иллюстрирующих работу симулятора.

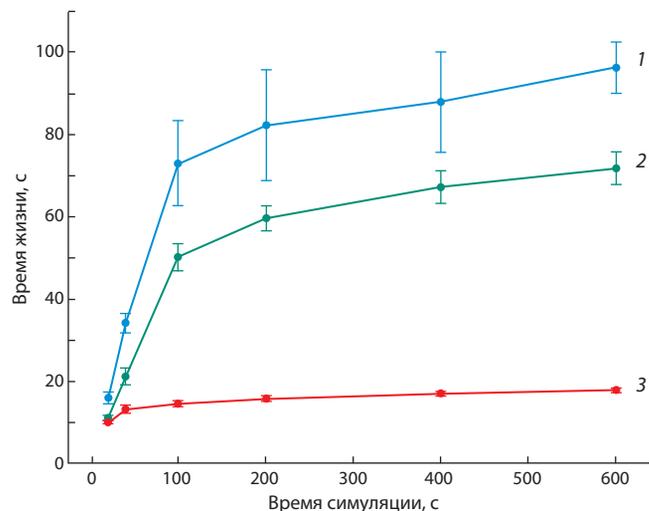
Работа генетического алгоритма может быть проиллюстрирована зависимостью времени жизни особей, возрастающей по мере роста числа поколений. Графики, представленные на рис. 4, были получены на основе 10 запусков симулятора с одинаковыми параметрами.

Видно, что с течением времени в популяции появляются особи, живущие во много раз дольше, чем те, которые еще не прошли естественный отбор (с параметрами нейросети, сгенерированными случайным образом). На уровне поведения и при визуальном наблюдении это выражается в том, что наиболее приспособленные виртуальные организмы целенаправленно двигаются к частицам «пищи» и избегают удаления от центральной области пространства, в которой «пища» сосредоточена, т. е. действительно являются успешно адаптировавшимися к своим условиям обитания.

## Обсуждение

Использованная нами архитектура нейронной сети достаточно проста и на данном этапе применялась в основном для тестирования системы в целом и для оценок ее производительности на ранней стадии разработки. В настоящее время осуществляется реализация следующей, значительно более продвинутой и современной архитектуры нейронной сети, представляющей собой комбинацию свёрточной нейронной сети (LeCun, Bengio, 1995) (для работы с входящими изображениями) и алгоритма NEAT (NEuroevolution of Augmenting Topologies) (Stanley, Miikkulainen, 2002). NEAT может изменять не только весовые параметры, но и структуру нейронной сети в процессе жизни организма. Свёрточная нейронная сеть будет преобразовывать детали изображения к некоторым абстракциям, а алгоритм NEAT – отвечать за поведенческую часть виртуального организма, работая с результатами функционирования свёрточной нейронной сети.

Помимо этого варианта, достаточно перспективными архитектурами являются самоорганизующиеся сети, например такие, как неокогнитрон (Kuniyiko, 1980). Также существуют значительно более реалистичные с точки зрения электрофизиологии нейронные сети, основанные на модели Ходжкина–Хаксли (Hodgkin, Huxley, 1952), в которой нервная клетка представлена в виде компартиментов, характеризующихся электрическими емкостями и сопротивлениями, с расчетами мембранных потенциалов и ионных токов. Современная реализация этой модели с поддержкой параллельных вычислений на GPU имеет следующие показатели производительности. В работе (Stimberg et al., 2020) нейросеть из 64 тыс. нейронов требует около 0.6 с вычислений на GPU Tesla V100 (обладает производительностью 14.1 TFLOPS в режиме FP32) для расчета 1 с времени симуляции (т. е. в режиме реального



**Рис. 4.** Зависимость лучшего времени жизни особи в популяции за весь период от начала до данного момента (1), на данный момент (2), а также среднее время жизни в популяции (3), с указанием среднеквадратичного отклонения.

Данные получены по 10 запускам симуляции.

времени), а нейросеть из 256 тыс. нейронов – около 3 с вычислений на 1 с времени симуляции. При этом численное интегрирование уравнений, описывающих систему, происходит с интервалом, не превышающим 0.1 мс для обеспечения точности вычислений и стабильности работы системы, а каждый нейрон в среднем имеет около 1000 связей, 80 % из которых активирующие, а 20 % – ингибирующие.

В последнее время достаточно активно ведутся исследования новых архитектур нейросетей, многие из которых находят практическое применение. В частности, в области нейроэволюционных методов накоплен широкий спектр перспективных вариантов, детально рассмотренный в работе (Хлопкова, 2016, гл. 1), включая их сравнение и классификацию, и в обзорной статье (Ma, Xie, 2022). В дальнейшем мы планируем реализовать наиболее подходящие и перспективные из них в представленном симуляторе и исследовать пределы их «когнитивных возможностей» при управлении моделью муравья.

## Заключение

Современные GPU, такие как, например, NVidia 3080 Ti, с 10240 параллельными вычислительными ядрами CUDA, обладают производительностью 34.1 TFLOPS, а ожидаемые в ближайшем будущем 4080 Ti – 67.6 TFLOPS. Таким образом, уже сейчас достигнута технологическая возможность осуществлять симуляцию одного виртуального организма с биологически реалистичной нейронной сетью из 256 тыс. нейронов и 256 млн связей между ними, с шагом численного интегрирования по времени 0.1 мс, на одном GPU. Это сравнимо с нейронной сетью нервной системы муравья, включающей около 250 тыс. нейронов.

Наши расчеты для виртуальных организмов с нейросетями из нескольких тысяч элементов показали, что вычислительные затраты на работу нейронных сетей и

виртуальную физическую среду относительно небольшие, а основным ограничивающим фактором являются потоки видеоданных в режиме «от первого лица», несущие зрительную информацию. Однако в случае нейросетей, состоящих из сотен тысяч нейронов, основным потребителем вычислительных ресурсов становится «нервная система». Таким образом, учитывая вышесказанное, современная настольная (desktop) вычислительная система с мощной современной видеокартой (GPU) обладает достаточной производительностью, чтобы обеспечить симуляцию виртуального организма с «нервной системой» на основе модели Ходжкина–Хаксли, эквивалентной по числу нейронов нервной системе муравья, в режиме реального времени. При наличии нескольких GPU в составе одного компьютера число одновременно моделируемых муравьев, взаимодействующих друг с другом, может быть увеличено пропорционально числу GPU.

## Список литературы / References

- Розенблатт Ф. Принципы нейродинамики. Перцептроны и теория механизмов мозга. М.: Мир, 1965.  
[Rosenblatt F. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Washington DC: Spartan Books, 1962.]
- Хлопкова О.А. Методы и алгоритмы интеллектуализации принятия решений в условиях неопределенности на базе аппарата нейронных сетей и эволюционного моделирования: Дис. ... канд. техн. наук. М., 2016.  
[Khlopkova O.A. Methods and algorithms for the intellectualization of decision making under conditions of indeterminacy based on neural networks and evolutionary modeling. Ph.D. Thesis. Moscow, 2016. (in Russian)]
- Aksoy V., Camlitepe Y. Spectral sensitivities of ants – a review. *Anim. Biol.* 2018;68(1):55-73. DOI 10.1163/15707563-17000119.
- Buehlmann C., Wozniak B., Goulard R., Webb B., Graham P., Niven J.E. Mushroom bodies are required for learned visual navigation, but not for innate visual behavior, in ants. *Curr. Biol.* 2020;30(17):3438-3443.e2. DOI 10.1016/j.cub.2020.07.013.
- Cammaerts M.-C., Cammaerts R. Are ants (Hymenoptera, Formicidae) capable of self recognition? *J. Sci.* 2015;5(7):521-532.
- Ferrario A., Palyanov A., Koutsikou S., Li W., Soffe S., Roberts A., Borisyuk R. From decision to action: detailed modelling of frog tadpoles reveals neuronal mechanisms of decision-making and reproduces unpredictable swimming movements in response to sensory signals. *PLoS Comput. Biol.* 2021;17(12):e1009654. DOI 10.1371/journal.pcbi.1009654.
- Herculano-Houzel S., Lent R. Isotropic fractionator: a simple, rapid method for the quantification of total cell and neuron numbers in the brain. *J. Neurosci.* 2005;25(10):2518-2521. DOI 10.1523/jneurosci.4526-04.2005.
- Herculano-Houzel S., Mota B., Lent R. Cellular scaling rules for rodent brains. *Proc. Natl. Acad. Sci. USA.* 2006;103(32):12138-12143. DOI 10.1073/pnas.0604911103.
- Hodgkin A.L., Huxley A.F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 1952;117(4):500-544. DOI 10.1113/jphysiol.1952.sp004764.
- Kunihiko F. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybernetics.* 1980;36(4):193-202. DOI 10.1007/BF00344251.
- LeCun Y., Bengio Y. Convolutional networks for images, speech, and time series. In: Arbib M.A. (Ed.) *The Handbook of Brain Theory and Neural Networks*. Cambridge; London: Bradford Book, The MIT press, 1995;276-278.
- Ma Y., Xie Y. Evolutionary neural networks for deep learning: a review. *Int. J. Mach. Learn. Cyber.* 2022;13:3001-3018. DOI 10.1007/s13042-022-01578-8.
- Macquart D., Garnier L., Combe M., Beugnon G. Ant navigation en route to the goal: signature routes facilitate way-finding of *Gigantops destructor*. *J. Comp. Physiol. A. Neuroethol. Sens. Neural. Behav. Physiol.* 2006;192(3):221-234. DOI 10.1007/s00359-005-0064-7.
- Moffett M.W., Garnier S., Eisenhardt K.M., Furr N.R., Warglien M., Sartoris C., Ocasio W., Knudsen T., Bach L.A., Offenberg J. Ant colonies: building complex organizations with minuscule brains and no leaders. *J. Org. Design.* 2021;10:55-74. DOI 10.1007/s41469-021-00093-4.
- Penick C.A., Ghaninia M., Haight K.L., Opachaloemphan C., Yan H., Reinberg D., Liebig J. Reversible plasticity in brain size, behavior and physiology characterizes caste transitions in a socially flexible ant (*Harpegnathos saltator*). *Proc. R. Soc. B. Biol. Sci.* 2021;288(1948):20210141. DOI 10.1098/rspb.2021.0141.
- Sarma G.P., Lee C.W., Portegys T., Ghayoomie V., Jacobs T., Alicea B., Cantarelli M., Currie M., Gerkin R.C., Gingell S., Gleeson P., Gordon R., Hasani R.M., Idili G., Khayrulin S., Lung D., Palyanov A., Watts M., Larson S.D. OpenWorm: overview and recent advances in integrative biological simulation of *Caenorhabditis elegans*. *Philos. Trans. R. Soc. B. Biol. Sci.* 2018;373(1758):20170382. DOI 10.1098/rstb.2017.0382.
- Scheffer L.K., Xu C.S., Januszewski M., Lu Z., Takemura S.Y., Hayworth K.J., Huang G.B., ... Meinertzhagen I.A., Rubin G.M., Hess H.F., Jain V., Plaza S.M. A connectome and analysis of the adult *Drosophila* central brain. *eLife.* 2020;9:e57443. DOI 10.7554/eLife.57443.
- Stanley K.O., Miikkulainen R. Evolving neural networks through augmenting topologies. *Evol. Comput.* 2002;10(2):99-127. DOI 10.1162/106365602320169811.
- Stimberg M., Goodman D.F.M., Nowotny T. Brian2GeNN: accelerating spiking neural network simulations with graphics hardware. *Sci. Rep.* 2020;10(1):410. DOI 10.1038/s41598-019-54957-7.

## ORCID ID

A. Palyanov [orcid.org/0000-0003-1108-1486](https://orcid.org/0000-0003-1108-1486)

**Благодарности.** Работа выполнена в рамках госзадания Института систем информатики им. А.П. Ершова СО РАН (тема FWNU-2022-0006).

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 19.08.2022. После доработки 01.11.2022. Принята к публикации 08.11.2022.

Перевод на английский язык <https://vavilov.elpub.ru/jour>

## Фосфолипазы A2 человека: функциональный и эволюционный анализ

И.И. Турнаев<sup>1,3</sup>✉, М.Е. Бочарникова<sup>2,3</sup>, Д.А. Афонников<sup>1,2,3</sup>

<sup>1</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>2</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

<sup>3</sup> Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия

✉ turn@bionet.nsc.ru

**Аннотация.** Фосфолипазы A2 (PLA2) способны гидролизовать *sn*-2 положение глицерофосфолипидов для высвобождения жирных кислот и лизофосфолипидов. Ферменты семейства фосфолипазы A2 широко распространены и присутствуют в большинстве клеток и тканей млекопитающих, выполняя функции регулятора метаболизма, поддержания мембранного гомеостаза, производства липидных медиаторов, ремоделирования мембран, активации воспалительных реакций. Соответственно, нарушение PLA2-регулируемого липидного метаболизма часто приводит к различным заболеваниям. В настоящем исследовании были систематически собраны и описаны 29 генов PLA2 в геноме человека на основе анализа литературных данных и изучения последовательностей. Анализ локализации генов PLA2 в геноме человека показал, что они расположены на 12 хромосомах человека и некоторые из них образуют кластеры. Оценка значений величины RVIS (оценка толерантности генов к мутациям, которые накапливаются в популяции человека) демонстрирует, что гены фосфолипаз A2 типа G4, входящие в один из двух наиболее крупных кластеров (четыре гена), наиболее толерантны к мутациям. Напротив, пониженную толерантность к мутациям имеют локализованные вне кластеров гены, кодирующие фосфолипазы A2 типа G6 (фосфолипазы A2 G6B, G6F, G6C, G6A). Мы проанализировали также связи между фосфолипазами A2 и заболеваниями человека по литературным данным, в результате чего выявлены связи 24 генов PLA2 со 119 заболеваниями, относящимися к 18 группам. Описано 229 связей «болезнь–ген» фосфолипазы A2. Показано, что белки фосфолипаз A2 типов G4, G2 и G7 вовлечены в наибольшее число заболеваний по сравнению с другими типами PLA2. С наибольшим числом типов PLA2 были связаны три группы заболеваний: новообразования, болезни системы кровообращения и болезни эндокринной системы. Филогенетический анализ показал, что общее происхождение устанавливается только для секреторных PLA2 (G1, G2, G3, G5, G10 и G12). Остальные типы PLA2 (G4, G6, G7, G8, G15 и G16) можно считать эволюционно независимыми. В результате проведенного анализа установлено, что наиболее толерантные к мутациям фосфолипазы A2 у человека (типы G4, G2 и G7) вовлечены в наибольшее количество групп заболеваний.

Ключевые слова: фосфолипаза A2; глицерофосфолипиды; заболевания человека.

**Для цитирования:** Турнаев И.И., Бочарникова М.Е., Афонников Д.А. Фосфолипазы A2 человека: функциональный и эволюционный анализ. *Вавиловский журнал генетики и селекции*. 2022;26(8):787-797. DOI 10.18699/VJGB-22-95

## Human phospholipases A2: a functional and evolutionary analysis

I.I. Turnaev<sup>1,3</sup>✉, M.E. Bocharnikova<sup>2,3</sup>, D.A. Afonnikov<sup>1,2,3</sup>

<sup>1</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup> Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia

✉ turn@bionet.nsc.ru

**Abstract.** Phospholipases A2 (PLA2) are capable of hydrolyzing the *sn*-2 position of glycerophospholipids to release fatty acids and lysophospholipids. The PLA2 superfamily enzymes are widespread and present in most mammalian cells and tissues, regulating metabolism, remodeling the membrane and maintaining its homeostasis, producing lipid mediators and activating inflammatory reactions, so disruption of PLA2-regulated lipid metabolism often leads to various diseases. In this study, 29 PLA2 genes in the human genome were systematically collected and described based on literature and sequence analyses. Localization of the PLA2 genes in human genome showed they are placed on 12 human chromosomes, some of them forming clusters. Their RVI scores estimating gene tolerance to the mutations that accumulate in the human population demonstrated that the G4-type PLA2 genes belonging to one of the two largest clusters (4 genes) were most tolerant. On the contrary, the genes encoding G6-type PLA2s (G6B, G6F, G6C, G6A) localized outside the clusters had a reduced tolerance to mutations. Analysis of the association between PLA2 genes and human diseases found in the literature showed 24 such genes were associated with 119 diseases belonging to 18 groups, so in total 229 disease/PLA2 gene relationships were described to reveal that G4, G2 and G7-type PLA2 pro-

teins were involved in the largest number of diseases if compared to other PLA2 types. Three groups of diseases turned out to be associated with the greatest number of PLA2 types: neoplasms, circulatory and endocrine system diseases. Phylogenetic analysis showed that a common origin can be established only for secretory PLA2s (G1, G2, G3, G5, G10 and G12). The remaining PLA2 types (G4, G6, G7, G8, G15 and G16) could be considered evolutionarily independent. Our study has found that the genes most tolerant to PLA2 mutations in humans (G4, G2, and G7 types) belong to the largest number of disease groups.

**Key words:** phospholipase A2; glycerophospholipids; human diseases.

**For citation:** Turnaev I.I., Bocharnikova M.E., Afonnikov D.A. Human phospholipases A2: a functional and evolutionary analysis. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):787-797. DOI 10.18699/VJGB-22-95

## Введение

Фосфолипазы (КФ 3.1) – это гидролазы, ферменты, которые используют молекулу воды для расщепления фосфолипидов (Burke, Dennis, 2009; Aloulou et al., 2018), основного компонента биологических мембран всех живых организмов (De Maria et al., 2007). Существует четыре класса фосфолипаз: А, В, С, D. Каждый из них способен гидролизовать определенную связь в фосфолипиде. Так, фосфолипазы А1 (PLA1) (КФ 3.1.1.32) и А2 (PLA2) (КФ 3.1.1.4) являются ацил-эстеразами и гидролизуют *sn*-1 и *sn*-2 позиции глицерофосфолипидов соответственно; фосфолипаза В (PLB) (КФ 3.1.1.5) гидролизует и *sn*-1, и *sn*-2 позиции глицерофосфолипидов; фосфолипазы С (PLC) (КФ 3.1.4.3) и D (PLD) (КФ 3.1.4.4) относятся к фосфат-эстеразам и определяются на основе гидролиза глицерина или дистальной стороны фосфатной группы (рис. 1) (Aloulou et al., 2018; Shayman, Tesmer, 2019).

Наиболее интенсивно изучают семейство фосфолипаз А2, что отражает их биологическую значимость. PLA2 гидролизуют эфирную связь мембранных фосфолипидов в позиции *sn*-2. В естественных условиях в *sn*-2 положении фосфолипидов часто содержится полиненасыщенные жирные кислоты, которые при высвобождении могут метаболизироваться с образованием различных эйкозаноидов и связанных с ними биологически активных липидных медиаторов (Aloulou et al., 2018).

Существует по меньшей мере шестнадцать типов фосфолипаз А2. Е.А. Dennis с коллегами разделили их на шесть групп, основываясь на свойствах: секреторируемые фосфолипазы А2 (sPLA2, типы G1, G2, G3, G5, G9, G10, G11, G12, G13 и G14); цитозольные фосфолипазы А2 (cPLA2, тип G4); кальций-независимые фосфолипазы А2

(iPLA2, тип G6); ацетилгидролазы PAF (platelet-activating factor – фактор активации тромбоцитов) (PAF-АН PLA2s, типы G7 и G8); лизосомальная фосфолипаза А2 (LPLA2, тип G15); фосфолипаза адипоцитов А2 (AdPLA, тип G16) (Dennis et al., 2011).

Отнесение PLA2 к определенной группе (типу) основано на экспериментальном определении их каталитических механизмов, клеточной локализации, эволюционных и структурных особенностях. Отметим, что большинство этих липолитических ферментов не имеют структурного сходства и обладают различными регуляторными и каталитическими механизмами (Aloulou et al., 2018).

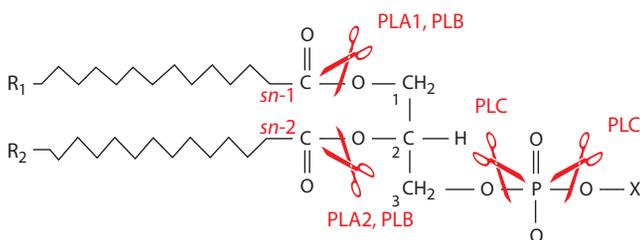
Каждый из шестнадцати типов фосфолипаз А2 вовлечен в различные виды липидного метаболизма и развития заболеваний, поэтому предполагается, что PLA2 могут служить терапевтическими мишенями для лечения ряда заболеваний (Aloulou et al., 2018). В связи с этим в фармацевтической промышленности наблюдается огромный интерес к разработке селективных и эффективных ингибиторов каждого из перечисленных типов PLA2 (Aloulou et al., 2018).

Известно, что описание функций белков включает, с одной стороны, молекулярную функцию, а с другой – функцию на уровне жизнедеятельности клетки или всего организма (Karp, 2000). Фосфолипазы А2 достаточно хорошо исследованы с точки зрения молекулярных функций и их механизмов. Однако их роль в процессах жизнедеятельности клетки и всего организма остается пока слабо изученной.

Настоящая работа посвящена анализу характеристик различных типов фосфолипаз А2 человека в контексте имеющихся данных о связях фосфолипаз А2 с разными заболеваниями. Были проанализированы: доменная организация белковых последовательностей, распределение генов фосфолипаз А2 в геноме, характеристики мутабельности генов фосфолипаз, а также их филогенетические отношения с фосфолипазами других организмов.

## Материалы и методы

**Формирование выборки фосфолипаз человека и животных.** Последовательности белков PLA2 человека были взяты из статьи (Huang Q. et al., 2015). Поскольку в этой работе описаны не все известные PLA2 человека (Dennis et al., 2011), то недостающие последовательности были идентифицированы в базе NCBI по их названиям и идентификаторам согласно (Dennis et al., 2011). Для этого использовали версию и аннотацию генома человека GRCh38.p14.



**Рис. 1.** Структурная схема фосфолипида и положения эфирных связей, гидролизующихся разными классами фосфолипаз.

R1 и R2 –  $((\text{CH}_2)_n-\text{CH}_3)$ ; X – различные полярные головные группы глицерофосфолипидов, например серин, холин, этаноламин, глицерол или инозитол; *sn*-1 и *sn*-2 – позиции в глицерофосфолипидах. Адаптировано из (Giresha, 2021).

Геном человека содержит 29 генов фосфолипаз А2, кодирующих белки двенадцати типов (PLA2G1–8, 10, 12, 15, 16) (последовательности см. в Прил. 2)<sup>1</sup>. Типы фосфолипаз А2 G1, G3, G5, G10, G15, G16 представлены одним геном, типы G7 (*G7A*, *G7B*), G8 (*G8A*, *G8B*) и G12 (*G12A*, *G12B*) – двумя генами, тип G2 – пятью генами (*G2A*, *G2C–F*), тип G6 – шестью генами (*G6A–F*).

Первичные структуры фосфолипаз А2 человека были охарактеризованы по наличию доменов, активных центров, сигнальных пептидов на основе литературных данных. Поиск гомологов фосфолипаз А2 человека у животных проводили в программе BLASTP с E-value ≤ 1. В качестве запроса использовали последовательности белков PLA2 человека. Гомологи искали среди белковых последовательностей организмов, представляющих различные таксоны, список которых приведен в Прил. 1.

**Функциональный анализ фосфолипаз.** Для того чтобы оценить консервативность генов фосфолипаз А2 человека, мы использовали метод RVIS (Petrovski et al., 2013). Индекс RVIS (residual variance index score) позволяет оценить толерантность генов к мутациям, которые накапливаются в популяции человека. Рассчитывается на основе информации о частоте аллелей, представленной в последовательности всего экзона человека (набор данных NHLBI-ESP6500 с сервера вариантов экзотов – EVS v.0.0.14: <https://evs.gs.washington.edu/EVS/>). RVIS позволяет ранжировать гены по количеству наблюдаемых нуклеотидных вариаций; при этом учитывается относительная доля нейтральных замен, которые наблюдаются для исследуемого гена. Отрицательные значения индекса RVIS свидетельствуют о низкой вариации гена (его последовательность является менее толерантной к накоплению мутаций, что обнаруживается у генов с более важной функцией), а положительные – о более высокой вариабельности гена (он более толерантен к заменам нуклеотидов).

Для выявления биологических процессов, в которые вовлечены фосфолипазы А2, мы использовали сервис DAVID (Huang D.W. et al., 2009). Он позволяет выявить термины из баз Gene Ontology, INTERPRO и KEGG Pathway, которые перепредставлены в аннотации генов из анализируемой выборки в сравнении с аннотациями всех генов организма. В нашем случае такой выборкой была выборка генов фосфолипаз А2 человека.

**Поиск ассоциации фосфолипаз А2 человека с заболеваниями.** Поиск статей, в которых описана связь между заболеваниями человека и активностью белков PLA2, проводился по базам PubMed и Google Scholar. Использовались запросы «disease/patients/pathology/название конкретного заболевания (например, lung cancer или schizophrenia) + PLA2/phospholipase A2/название конкретной PLA2 (например, pla2g1b, pla2g2a)». Также информация была взята из обзоров, посвященных участию фосфолипаз А2 в различных группах заболеваний.

В статьях, полученных в результате поиска, отслеживали информацию о связи заболевания человека и активности/экспрессии определенного PLA2. К такой информации, например, относились сообщения о том, что у больных значительно понижена/повышена экспрессия

или активность данного PLA2 по сравнению со здоровыми людьми; сведения о том, что мутация гена PLA2 усиливает/ослабляет выраженность заболевания, и о том, что установлен механизм, посредством которого PLA2 влияет на ход заболевания. Для классификации болезней в настоящем исследовании использовали Международную классификацию болезней МКБ-10 (<https://mkb-10.com>) (Hirsch et al., 2016).

На основании информации о взаимосвязи между заболеванием и вовлечением в него фосфолипаз различного типа была сформирована таблица данных, по строкам которой расположены типы заболеваний человека, а по столбцам – типы фосфолипаз А2. Если в ячейке таблицы стояло значение 1, это означало, что данный тип фосфолипазы вовлечен/имеет связь с заболеванием. Для построения таблицы был написан скрипт на языке Python, который осуществлял в процессе работы привязку названия заболевания к коду МКБ.

Далее мы проводили иерархическую кластеризацию типов фосфолипаз А2 человека по степени их ассоциации с различными заболеваниями: для разных типов фосфолипаз сравнивали значения их участия в заболевании из вышеописанной таблицы. В качестве меры сходства использовали евклидово расстояние. Для кластеризации применяли метод UPGMA. Аналогично кластеризовали и заболевания, по степени их связи с различными типами фосфолипаз А2.

**Множественное выравнивание и реконструкция филогении белков.** Множественное выравнивание гомологичных последовательностей фосфолипаз осуществляли с помощью программ PROMALS (Pei, Grishin, 2007) и MAFFT (Katoh, Toh, 2010). Поиск белков для выравнивания и, соответственно, само выравнивание белковых последовательностей проводились только по фосфолипазному домену. Реконструкцию филогенетического дерева выполняли методом максимального правдоподобия (программа IQ-TREE v.8.2.4) (Nguyen, 2015). Для реконструкции дерева была выбрана оптимальная модель WAG + R6.

## Результаты

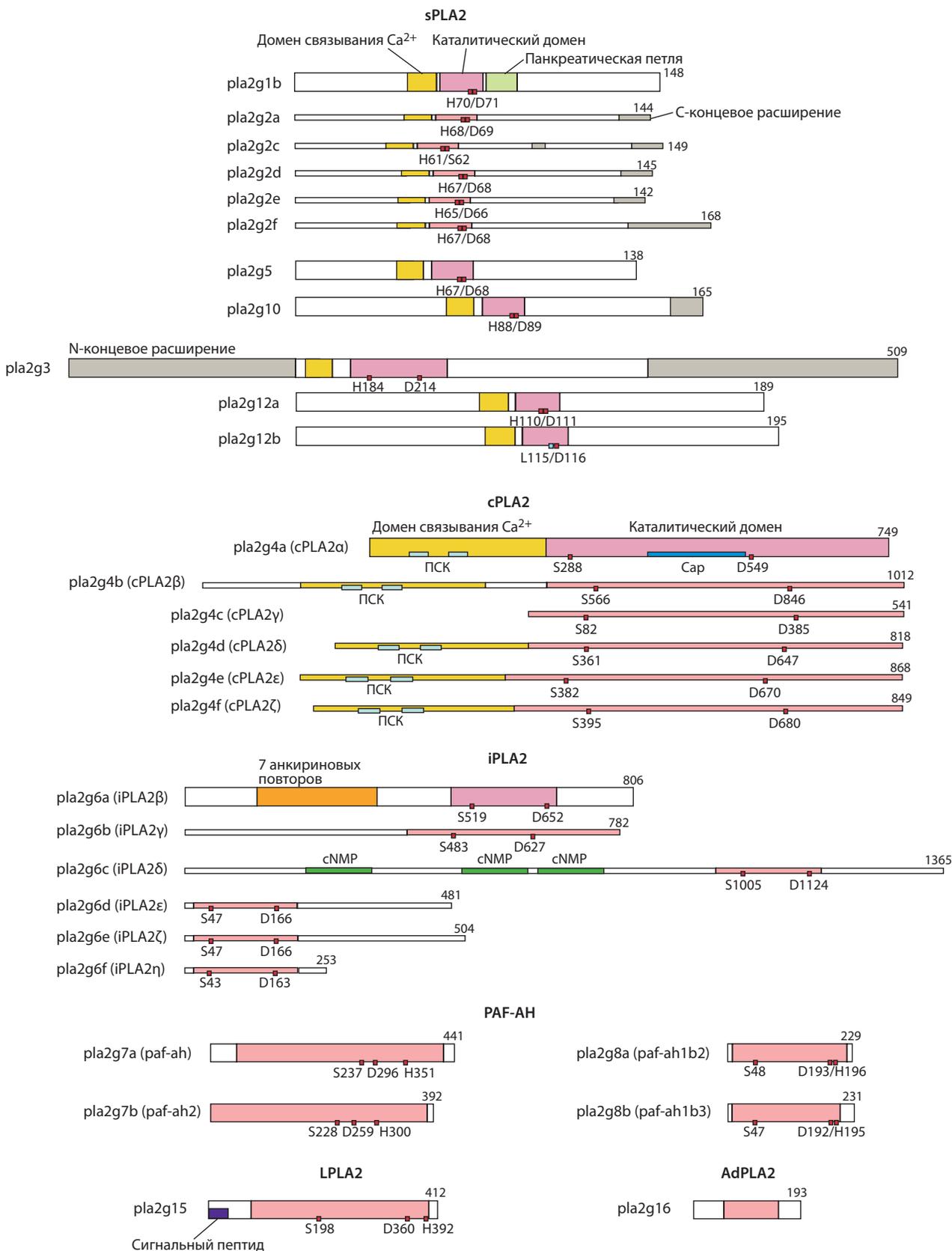
### Структурная и функциональная характеристика фосфолипаз человека

Особенности структурной организации фосфолипаз человека различных типов отражены на рис. 2. Характеристика белков (субстраты, активность, масса, каталитические остатки и т. д.) приведена в Прил. 3.

**Секреторные фосфолипазы А2 (sPLA2)** включают шесть типов PLA2: G1, G2 (pla2g2(a, c–f)), G3, G5, G10, G12 (pla2g12(a, b)). Длина белков типов G1, G2, G5, G10 составляет 138–165 а. о., типа G12 – 189–195 а. о., тогда как белок типа G3 значительно крупнее – 509 а. о. (см. рис. 2). Его размер значительно увеличен по сравнению с остальными sPLA2 белками за счет протяженных С- и N-концевых расширений.

Для белков sPLA2 (типы G1, G2, G3, G5, G10, G12) показана только PLA2 каталитическая активность. Каталитические домены расположены в центральной части последовательностей и содержат консервативную ката-

<sup>1</sup> Приложения 1–11 см. по адресу: <https://sites.icgbio.ru/vogis/download/pict-2022-26/appx13.pdf>



**Рис. 2.** Структура белков PLA2 человека.

Красным квадратиком обозначены активные сайты. Голубой квадратик в последовательности pla2g12b обозначает замену H (гистидина) на L (лейцин) в 115-й позиции белка, что приводит к отсутствию каталитической активности белка pla2g12b (Guan et al., 2011). ПСК – петля, связывающая Ca<sup>2+</sup>; cNMP – домен, связывающий циклические нуклеотиды (цАМФ или цГМФ); панкреатическая петля – уникальное пятиаминокислотное расширение, имеющееся у sPLA2G1B; Cap – область крышки, обнаруженная в PLA2G4A, которая закрывает/открывает активный сайт для моделирования фосфолипидного субстрата. Рисунок адаптирован из статей (Kudo, Murakami, 2002; Dennis et al., 2011).

литическую диаду His/Asp (см. рис. 2, Прил. 3). sPLA2 являются  $Ca^{2+}$ -зависимыми фосфолипазами, и, соответственно, у них ближе к N-концу есть домен связывания кальция, включающий в себя петлю связывания кальция (Dennis et al., 2011) (см. рис. 2).

Субстратами для ферментов sPLA2 служат фосфолипиды. Во всех случаях это фосфатидилхолин (ФХ) и фосфатидэтаноламин (ФЕ), за исключением pla2g12a, который использует фосфатидилглицерол (ФГ), но не ФХ или ФЕ. Некоторые sPLA2 используют в качестве субстратов ФГ и фосфатидилсерин (ФС). Белок pla2g12b человека каталитически неактивен (см. Прил. 3 и подпись к рис. 2).

**Цитозольные фосфолипазы А2 (сPLA2)** представлены типом G4. К этому типу относятся шесть белков человека: pla2g4(a–f). Масса белков сPLA2 варьирует от 541 а.о. (pla2g4c) до 1012 а.о. (pla2g4b) (см. рис. 2). У белков сPLA2 каталитические домены расположены в С-концевой части последовательностей и содержат консервативную каталитическую диаду Ser/Asp (см. Прил. 3, рис. 2). сPLA2, так же как и sPLA2, являются кальций-зависимыми фосфолипазами А2, и, соответственно, у них ближе к N-концу тоже есть домен связывания кальция (Dennis et al., 2011).

Для белков G4 типа (сPLA2/PLA2G4), как и для белков sPLA2, отмечена активность PLA2, в качестве субстратов для которой используются ФХ и ФЕ. Белок pla2g4a дополнительно задействует как субстрат фосфатидилинозитол (ФИ), а белок pla2g4c – ФХ, но его специфичность для ФЕ не показана (см. Прил. 3).

**Кальций-независимые фосфолипазы А2 (iPLA2)** включают только фосфолипазы А2 типа G6. Длина белков PLA2G6 человека сильно варьирует – от 253 до 1365 а.о. В белках iPLA2 каталитические домены расположены у pla2g(a, c) ближе к С-концу, а у pla2g(d–f) ближе к N-концу (см. рис. 2). Каталитический домен содержит, как и у белков сPLA2, консервативную каталитическую диаду Ser/Asp (см. Прил. 3, рис. 2). Каталитическая активность iPLA2 не зависит от присутствия  $Ca^{2+}$ , что отражено в их названии «кальций-независимые фосфолипазы». В отличие от sPLA2 и сPLA2, они не имеют  $Ca^{2+}$ -связывающего домена. У белка pla2g6a ближе к N-концу находится область, содержащая семь анкириновых повторов. Данный мотив участвует в белок-белковых взаимодействиях, позволяя интенсивно связываться с мембранными белками (Филькин и др., 2020). В белке pla2g6c ближе к С-концу находятся три cNMP сайта (сайт, связывающий циклические нуклеотиды) (см. рис. 2).

Белки типа g6(a–f) используют в качестве субстратов для PLA2 реакций ФХ. Белок pla2g6b, кроме ФХ, может задействовать ФЕ. Также эти ферменты, помимо PLA2, могут проявлять и другие активности: триглицерид гидролазную, лизофосфолипазную, PLA1 (для pla2g6b) и др. (см. Прил. 3).

**Ацетилгидролазы фактора активации тромбоцитов (РАФ-АН или Lp-PLA2).** РАФ-АН включают фосфолипазы А2 типов G7 и G8. Они моделируют активность фактора активации тромбоцитов. Фактор активации тромбоцитов (platelet-activating factor, PAF) – сильный фосфолипидный медиатор воспаления. PAF участвует в вос-

палении, агрегации тромбоцитов, играет роль в патогенезе анафилактического шока (Shimizu, 2009). Длина РАФ-АН белков составляет 441 и 392 а.о. для белков g7a и g7b соответственно и 229 и 231 а.о. для белков g8a и g8b (см. рис. 2). Каталитический домен занимает почти всю последовательность и содержит консервативную каталитическую триаду Ser/His/Asp (см. Прил. 3, рис. 2). Белки РАФ-АН являются фосфолипазами А2, не зависящими от  $Ca^{2+}$ . Соответственно, у них отсутствует домен связывания  $Ca^{2+}$  (см. рис. 2).

Белки pla2g7(a, b) и pla2g8(a, b) способны гидролизовать РАФ (фосфолипидный фактор активации тромбоцитов) до lysoPAF. При этом белок pla2g7a обладает как PLA2, так и PLA1 активностями и может использовать в качестве субстрата ФХ и оксФХ. Белок pla2g7b показывает PLA2 активность (см. Прил. 3).

**Лизосомальные фосфолипазы А2 (LPLA2).** LPLA2 (PLA2 тип G15) представлен у человека одним белком pla2g15 длиной 412 а.о. (см. рис. 2). У pla2g15 каталитический домен находится в центральном районе последовательности и содержит консервативную каталитическую триаду Ser/His/Asp (см. Прил. 3, рис. 2). Каталитическая активность PLA2 белка lpla2 не зависит от присутствия  $Ca^{2+}$ , поэтому последовательность lpla2 не имеет домена для связывания с  $Ca^{2+}$ .

Белок pla2g15/lpla2 типа G15 обладает PLA2 и PLA1 активностями, субстратами для которых служат ФХ, ФЕ и ФС. Также pla2g15 способен к ацилцерамид синтазной активности, субстратом для которой служит С1 церамид (см. Прил. 3).

**Фосфолипазы адипоцитов А2 (AdPLA2).** AdPLA2 (PLA2 тип G16) представлен у человека одним белком pla2g16 длиной 193 а.о. (см. рис. 2). Каталитический домен находится в центральном районе последовательности и содержит, как и у белков LPLA2, консервативную каталитическую триаду Ser/His/Asp (см. Прил. 3, рис. 2). adpla2 является  $Ca^{2+}$ -независимой фосфолипазой А2, и, соответственно, у него, как и у iPLA2, РАФ-АН, LPLA2, отсутствует домен для связывания  $Ca^{2+}$  (см. рис. 2).

Белок G16 типа pla2g16/adpla обладает PLA2 и PLA1 активностями, используя в качестве субстратов ФХ и ФЕ. Также он имеет N-ацил ФЕ ацилтрансферазную активность, задействуя как субстрат диацил-ФЕ (см. Прил. 3).

### Локализация генов фосфолипаз А2 в геноме человека

Расположение генов фосфолипаз А2 в геноме человека (версия GRCh38.p14) представлено на рис. 3. Гены PLA2 отсутствуют на хромосомах 2, 3, 5, 8, 9, 13, 14, 17, 18, 20, 21 и Y хромосоме. На хромосомах 4, 6, 7, 10, 12 и X локализовано по одному гену PLA2, на хромосомах 11, 16 – по два гена, а на хромосомах 19, 22 – по три. На хромосоме 15 четыре гена (G4B, G4E, G4D, G4F) образуют кластер размером 0.3 Мб в положении 43 Мб. На хромосоме 1, кроме одиночного гена G4A (на 188 Мб), в положении 20 Мб находится кластер из шести генов (G2E, G2A, G5, G2D, G2F, G2C) размером 0.11 Мб. Надо отметить, что все остальные гены, за исключением генов названных двух кластеров, находятся друг от друга на хромосомах изолированно, на расстоянии не менее 6 Мб друг от друга. При этом если

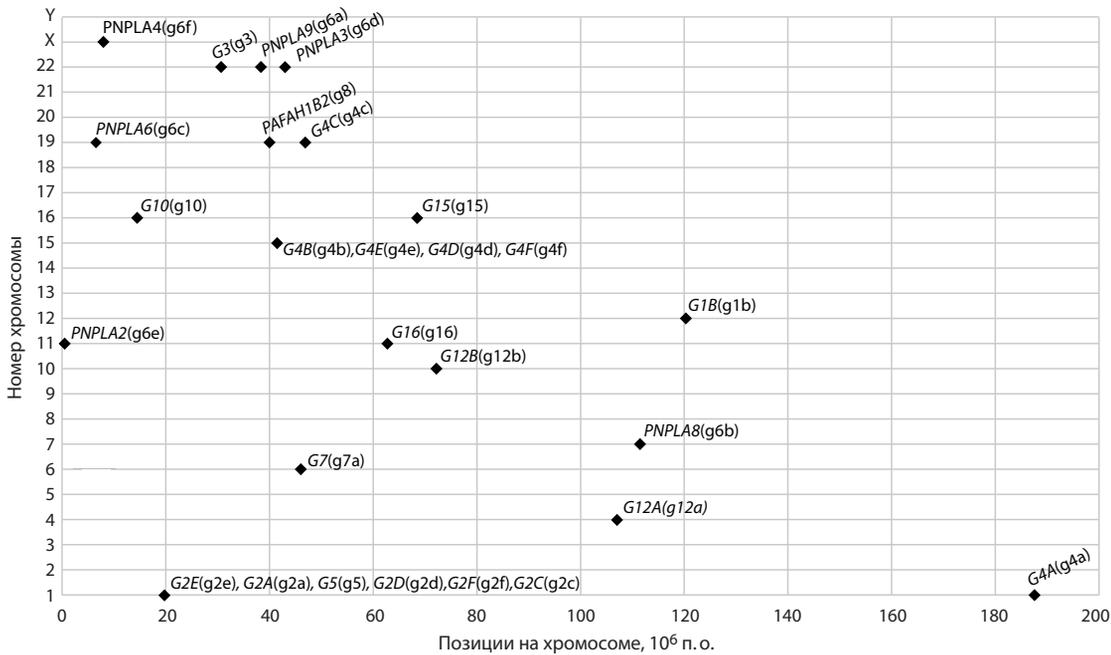


Рис. 3. Локализация генов фосфолипаз A2 на хромосомах человека.

Гены обозначены ромбами. Их позиции по оси X соответствуют координатам начала генов на хромосоме, по оси Y – номеру хромосомы.

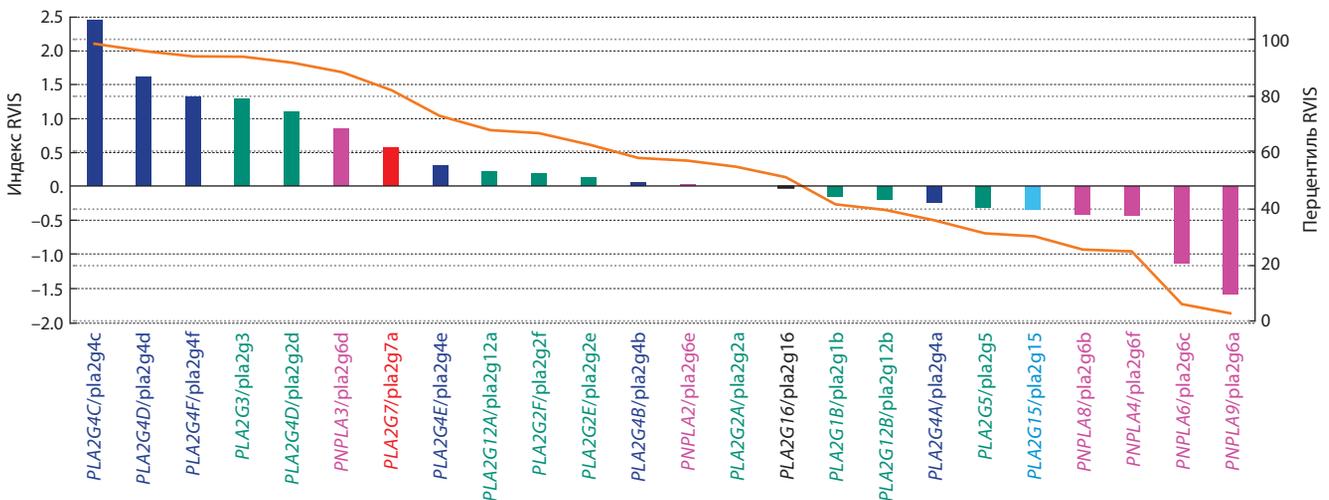


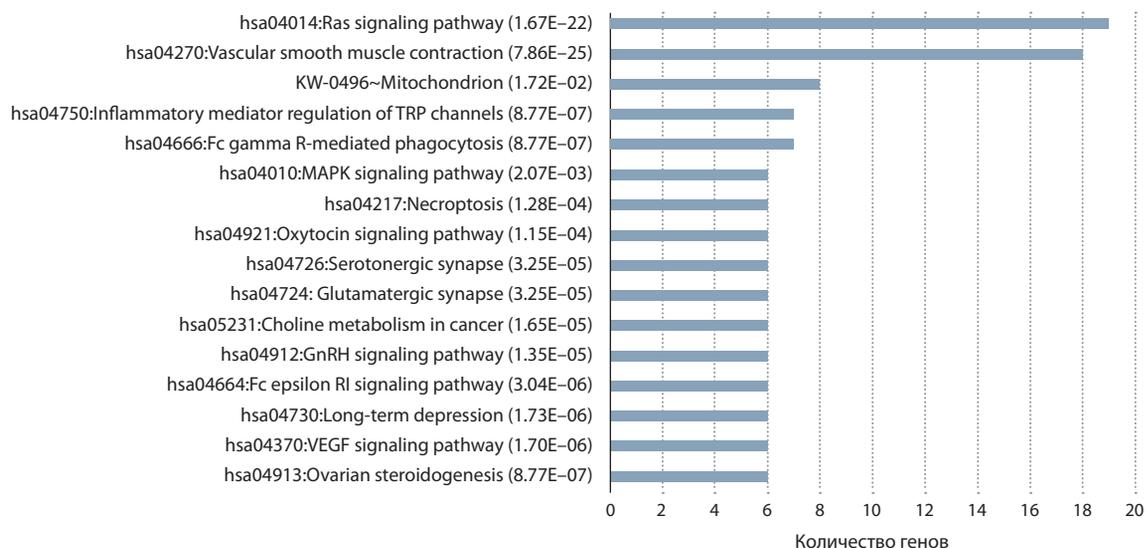
Рис. 4. Распределение значений индекса RVIS (Residual Variation Intolerance Score) для генов фосфолипаз A2 в геноме человека.

Ось Y справа соответствует доле генов в геноме человека (в %), имеющих меньшее значение RVIS, чем для конкретного гена (столбца). Этим значениям перцентилей соответствует оранжевая линия на графике. Цвет столбцов отражает принадлежность фосфолипаз A2 к разным типам, где синий – цитозольные PLA2, зеленый – секретируемые, розовый – кальций-независимые, красный – ацетилтрансферазы фактора активации тромбоцитов (PAF-AH), черный (pla2g16) – адипотические PLA2, голубой (pla2g15) – лизосомальные PLA2.

гены PLA2 типа G4 (*G4B*, *G4E*, *G4D*, *G4F*) локализованы в вышеописанном кластере на хромосоме 15, то два других гена этого типа (*G4A* и *G4C*) расположены изолированно на хромосомах 1 и 19 соответственно. При этом все гены PLA2 типа G2 (*G2A*, *G2C*, *G2D*, *G2E*, *G2F*) находятся в кластере на хромосоме 1, но вместе с ними в этот кластер входит ген *G5*. Гены фосфолипаз A2 типа G6 расположены: *G6B* – на хромосоме 7, *G6E* – на хромосоме 11, *G6C* – на хромосоме 19, *G6A* и *G6D* – на хромосоме 22, *G6F* – на X хромосоме.

### Анализ толерантности генов фосфолипаз A2 к мутациям на основе величины RVIS

На рис. 4 представлено распределение индексов RVIS для PLA2 человека. Слева на графике располагаются PLA2 с индексом выше нуля; это гены, которые содержат относительно большое число мутаций и толерантные к ним. Справа – фосфолипазы с индексом ниже нуля (гены, которые менее толерантны к мутациям). Негативное значение RVIS имеют гены фосфолипазы A2 типов G16, G1, G12 (*PLA2G12B*), G4 (*PLA2G4A*), G5, G15, G6 (см. рис. 4). Из



**Рис. 5.** Сигнальные пути и биологические процессы из базы KEGG Pathway, значимо ассоциированные со списком генов фосфолипаз А2 по данным анализа, проведенного с использованием системы DAVID.

По оси Y расположены термины, описывающие сигнальные пути и биологические процессы. В скобках после каждого термина приведено значение FDR (false discovery rate – ожидаемая доля ложных отклонений). По оси X отложено количество генов фосфолипаз А2, ассоциированных с каждым термином, описывающим сигнальные пути и биологические процессы.

этих генов три относятся к секретируемым PLA2 типов G1, G12, G5, а остальные относятся к кальций-независимым (тип G6), цитозольным (G4), лизосомальным (G15) и адипотическим (G16) фосфолипазам А2. Интересно, что четыре из шести генов PLA2 типа G6 (кальций-независимые PLA2), *PNPLA8*, *PNPLA4*, *PNPLA6*, *PLA2G6* (кодирующие белки *pla2g6b*, *pla2g6f*, *pla2g6c*, *pla2g6a*), имеют самые низкие величины показателя RVIS, т. е. наименее толерантны к мутациям. Остальные гены фосфолипаз этого типа (*PNPLA2* и *PNPLA3*, кодирующие белки *pla2g6e* и *pla2g6d*) имеют RVIS, свидетельствующий об умеренной или немного повышенной толерантности.

Наиболее толерантными к мутациям оказались гены фосфолипаз А2 цитозольного типа G4. Пять из шести генов этого типа имеют положительное значение RVIS, и только у гена *PLA2G4A* значение RVIS меньше 0 ( $RVIS = -0.25$ ). Привлекает внимание, что более толерантные к мутациям гены PLA2 типа G4 *PLA2G4(B, D-F)* локализованы в кластере на хромосоме 15, в отличие от гена *PLA2G4A* с низкой толерантностью к мутациям, который располагается отдельно от них, на хромосоме 1. Для пяти PLA2 человека (*PLA2G2C*, *PLA2G7B*, *PLA2G8A*, *PLA2G8B*, *PLA2G10*) на сервере EVS отсутствуют данные по вариабельности генов, по которым рассчитывался RVIS (см. Материалы и методы), в связи с чем этих PLA2 нет на графике (см. рис. 4).

#### Анализ связи генов фосфолипаз А2 человека с биологическими процессами и сигнальными путями из базы KEGG Pathway

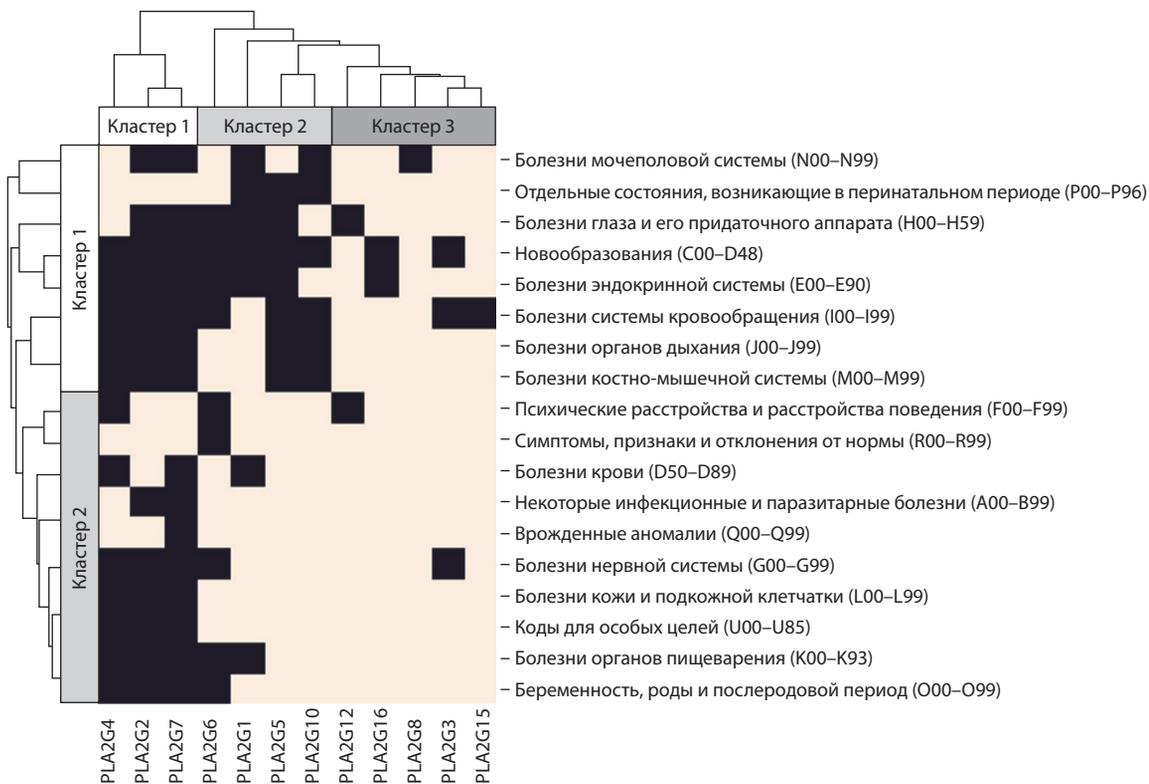
Результаты функционального анализа списка генов фосфолипаз А2 человека, выполненного при помощи сервера DAVID, приведены на рис. 5. Оказалось, что наиболее значимым (по количеству ассоциированных с ним генов фосфолипаз А2) является «сигнальный путь Ras». Ин-

тересно, что этот путь вовлечен в канцерогенез. Другой термин – путь VEGF, связанный с васкулярным эндотелиальным фактором роста. Заболевания, ассоциированные с данным путем, тоже, как правило, связаны с развитием опухолей (рак молочной железы, глиома, меланома и др.) (Takahashi, Shibuya, 2005). Таким образом, эти данные показывают значительную связь фосфолипаз А2 с процессами канцерогенеза.

#### Болезни, связанные с фосфолипазами А2

По информации из научных статей мы провели анализ связи между активностью различных фосфолипаз А2 и заболеваниями у человека (Прил. 4). Заболевания обозначены согласно Международной классификации болезней (МКБ-10, <https://mkb-10.com>). Отражена связь заболевания и PLA2 в случаях, если в статьях есть информация: 1) об ассоциации экспрессии определенного PLA2 с ходом заболевания; 2) о связи мутаций PLA2 с течением заболевания; 3) описаны механизмы воздействия PLA2 на ход заболевания. Описаны 229 связей «болезнь–ген», показаны связи 24 генов PLA2, входящих в 12 типов PLA2, со 119 заболеваниями (см. Прил. 4).

Результаты кластеризации фосфолипаз различного типа по их связи с заболеваниями человека приведены на рис. 6. По строкам диаграммы представлены 18 групп заболеваний, их названия и коды МКБ-10 (в скобках). Столбцы диаграммы соответствуют 12 типам PLA2 человека. Больше всего групп PLA2 связано с новообразованиями (код МКБ-10: C00–D48) – 9 групп из 12-ти; болезнями системы кровообращения (I00–I99) – 8; болезнями эндокринной системы (E00–E90) – 7; болезнями глаза и его придаточного аппарата (H00–H59) – 6 групп. Наименьшее количество групп PLA2 связано с врожденными аномалиями (Q00–Q99) – только одна PLA2 типа G7; с группой болезней «симптомы, признаки и отклонения от нормы»



**Рис. 6.** Связь фосфолипаз A2 с различными группами заболеваний.

Черные ячейки – выявлена связь между генами группы фосфолипаз A2 и заболеваниями из представленных групп заболеваний; светлые – связей «ген–заболевание» не выявлено.

(R00–R99) – только PLA2 типа G6; с группой болезней «некоторые инфекционные и паразитарные болезни» (A00–B99) – только PLA2 типов G2 и G7.

Интересно отметить, что из PLA2 больше всего связей с заболеваниями имеют следующие: тип G7 связан с 15 группами заболеваний из 18 представленных на рис. 6; G2 – с 13 группами; G4 – с 12 группами. Меньше всего представлены в группах заболеваний: PLA2 типа G8, связанные только с болезнями мочеполовой системы (N00–N99); типа G15 – только с болезнями системы кровообращения (I00–I99); типа G12 – только с психическими расстройствами и расстройствами поведения (F00–F99) и с болезнями глаза и его придаточного аппарата (H00–H59); типа G16 – только с новообразованиями (C00–D48) и с болезнями эндокринной системы (E00–E90).

При кластеризации фосфолипазы A2 (по горизонтали) разделились на три кластера (см. рис. 6). Первый кластер содержит фосфолипазы A2 типов G4, G2, G7 и характеризуется тем, что гены этих типов PLA2 участвуют в наибольшем числе проанализированных нами заболеваний человека. Второй кластер включает PLA2 типов G6, G5, G10, которые относятся к секретируемым фосфолипазам A2. Они участвуют примерно в половине из проанализированных нами заболеваний (болезни различных систем: мочеполовой (N00–N99), кровообращения (I00–I99), дыхания (J00–J99); новообразования (C00–D48) и т. д.). PLA2 третьего кластера (типы G12, G16, G8, G3, G15) можно считать специфическими по отношению к отдельным заболеваниям.

Заболевания также разделились на две большие группы: первый кластер – болезни, в которых участвуют фосфолипазы A2 большинства типов; второй кластер – заболевания, в которые вовлечены фосфолипазы A2 в основном из кластера 1 (кластер фосфолипаз A2) (см. рис. 6).

Следует отметить, что одиннадцать из двенадцати исследованных генов PLA2 типов G4, G2, G7 имеют высокий уровень толерантности к мутациям (RVIS), и только ген *PLA2G4A* – умеренно низкий уровень толерантности к мутациям (RVIS = –0.25) (см. рис. 4). При этом PLA2 этих типов (G4, G2, G7) участвуют в наибольшем числе заболеваний (см. рис. 6). В то же время из семи исследованных генов PLA2 типов G6 и G15 пять имеют самый низкий уровень толерантности к мутациям (RVIS), и только один, *PLA6G6D*, отличается относительно высоким уровнем толерантности к мутациям (RVIS = 0.85) (см. рис. 4). PLA2 этих типов (G6 и G15) относятся к кластеру фосфолипаз, связанных с наименьшим числом заболеваний (см. рис. 6). Это позволяет предполагать возможную положительную зависимость между количеством заболеваний, в которые вовлечена фосфолипаза A2, и толерантностью гена к мутациям (индекс RVIS). Однако расчет коэффициента корреляции методом  $\chi^2$  не выявил значимой корреляции между этими величинами. В данном случае можно говорить только о недостоверной тенденции.

В работе (Petrovski et al., 2013) в ходе исследования выборки генов, связанных с менделевскими (моногенными) заболеваниями, было достоверно показано, что они имеют низкую толерантность (индекс RVIS) по сравнению с

остальными генами человека. S. Petrovski с коллегами предполагают, что отрицательный RVIS свидетельствует о наличии очищающего (purifying) отбора, а положительный RVIS – либо об отсутствии очищающего отбора, либо даже о наличии некоторых форм балансирующего (balanced) или движущего (positive) отбора.

Тенденция к повышенному уровню индекса RVIS у генов фосфолипаз А2, вовлеченных в большее количество заболеваний, возможно, обусловлена тем, что при рассмотрении экспрессионных данных набор выявленных дифференциально экспрессирующихся генов был существенно зашумлен случайными генами. Появление таких случайных генов может быть связано как с особенностями методики (Hatfield et al., 2003), так и с тем, что любое возмущение в клетке и организме (например, заболевание) индуцирует неспецифические эффекты на экспрессию генов (например, активирует гены стрессового ответа, апоптоза, некроз и т.д.) (Leuner et al., 2007; Turkmen, 2017). Поэтому при оценке числа связей между фосфолипазами А2 и заболеваниями большое или малое количество связей необходимо интерпретировать с некоторой осторожностью.

### Эволюция фосфолипаз А2

С помощью поиска по гомологии были идентифицированы последовательности фосфолипаз А2 для 32 видов организмов (см. Прил. 1), включая 13 позвоночных и 19 беспозвоночных (см. Материалы и методы). Их идентификаторы и последовательности приведены в выборке Прил. 5.

Для иллюстрации сходства функциональных районов PLA2 мы провели анализ гомологии между каталитическими доменами фосфолипаз А2 у человека. Сходство между доменами PLA2 разных типов выявлено только среди секреторных PLA2 (Прил. 6), а именно: между каталитическими доменами PLA2 типов G1, G2, G5, G10 есть сходство (E-value) от  $2e-03$  до  $2e-38$ ; между белками типа G12 (g12a и g12b) E-value =  $4e-48$ . Нет сходства (E-value  $\geq 1$ ) между PLA2 типа G3 (plag3) и всеми остальными белками sPLA2.

Во всех остальных типах PLA2, кроме принадлежащих к sPLA, не отмечено сходства доменов между PLA2 разных типов. Внутри типов, в частности, сходство между цитозольными PLA2 (тип G4) варьировало (E-value) от  $4e-27$  до  $1e-177$  (Прил. 7).

Кальций-независимые PLA2 (тип G6) по сходству доменов распадаются на три подгруппы: 1) pla2g6d, 6e, 6f со сходством доменов от  $2e-42$  до  $4e-91$ ; 2) pla2g6a, 6b со сходством между их каталитическими доменами (E-value)  $4e-12$ ; 3) каталитический домен pla2g6c, не имеющий гомологии с каким-либо другим белком 6-го типа (Прил. 8). Соответственно, между доменами человеческих PLA2 этих трех подтипов типа G6 сходство отсутствовало. Типы PLA2 G7 (два белка g7a и g7b) и G8 (два белка g8a и g8b) имеют гомологию внутри типов 7e–103 (тип G7) (Прил. 9) и 3e–102 (тип G8) (Прил. 10). Для оставшихся двух типов сравнение не проводилось, так как они включают только по одному белку у человека.

Чтобы реконструировать филогению белков PLA2, мы предварительно провели анализ гомологии и множест-

венного выравнивания между белками различных типов PLA2. В результате определено, что белки, объединенные в группу «секретируемые sPLA2» (типы G1–3, G5, G10, G12), обладают высокой или умеренной гомологией (E-value  $\leq 1$ ) и качественно выравниваются между белками этих групп. Белки других типов PLA2 (G4, G6, G7, G8, G15, G16), напротив, обладают очень низкой гомологией (E-value  $> 1$ ) и плохо выравниваются между группами и по отношению к белкам sPLA2. Исходя из этого мы реконструировали филогению sPLA2 методом максимального правдоподобия (рис. 7). Для остальных типов белков (G4, G6, G7, G8, G15, G16) филогенетический анализ не проводился, так как реконструкция филогении по некачественному выравниванию с высокой вероятностью приведет к артефактной филогении.

Результаты анализа филогении sPLA2 предполагают, что у общих предков многоклеточных беспозвоночных животных произошло две последовательных дивергенции: сначала предковый ген sPLA2 дивергировал на гены G3/G12 и G1/G2/G5/G10, а затем ген G3/G12 дивергировал на предковые гены G3 и G12.

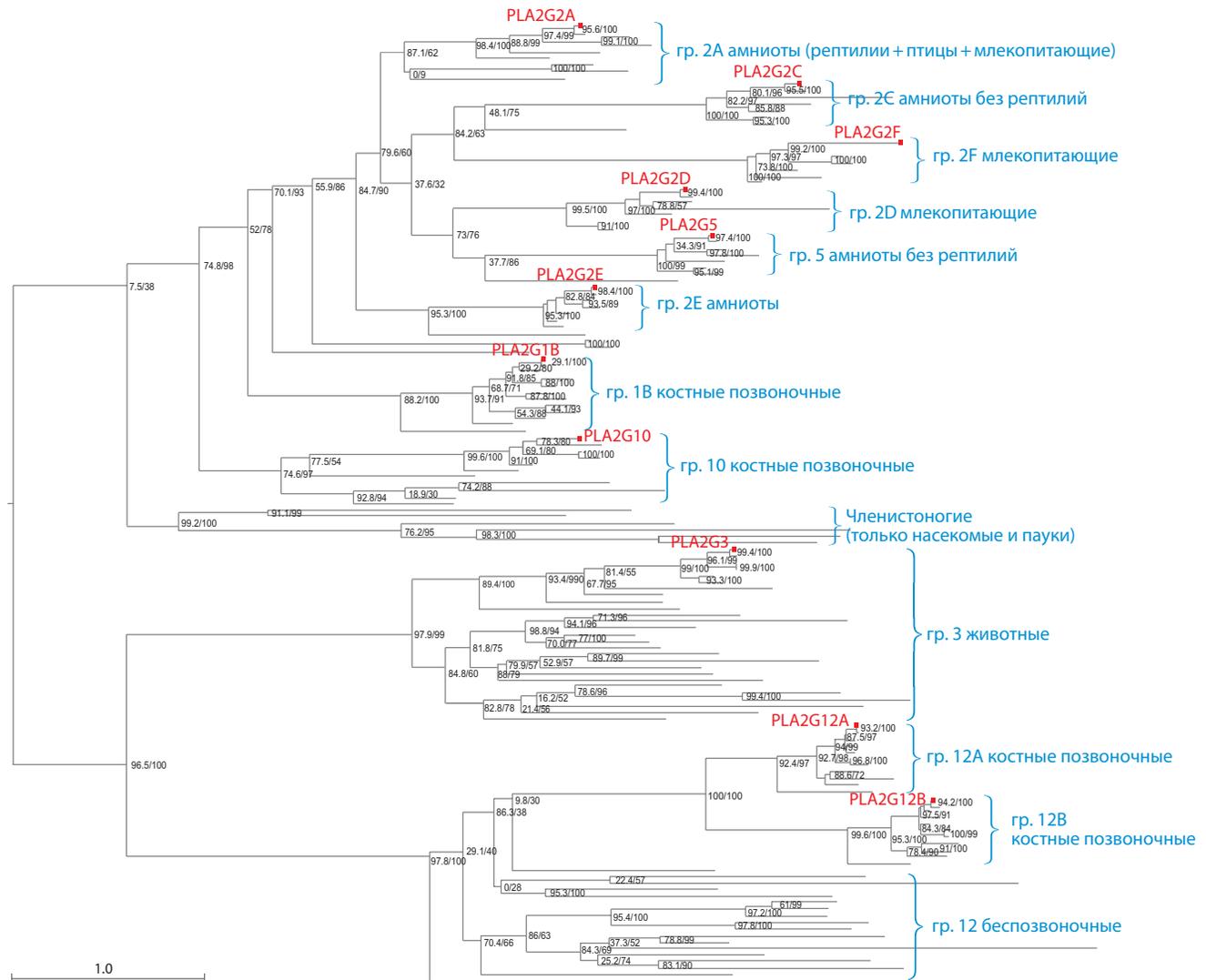
У общих предков костных позвоночных предковый ген G12 дивергировал на гены G12A и G12B, тогда как предковый ген G1/G2/G5/G10 дивергировал у общих предков костных позвоночных на гены G10 и G1/G2/G5, а потом ген G1/G2/G5 дивергировал на гены G1 и G2/G5. Далее у общих предков амниот произошли дивергенции гена G2/G5 на гены G2E и G2A/G2C/G2D/G2F/G5, затем гена G2A/G2C/G2D/G2F/G5 на гены G2A и G2C/G2D/G2F/G5, далее гена G2C/G2D/G2F/G5 на гены G2C/G2F и G2D/G5, и в итоге ген G2C/G2F дивергировал на гены G2C и G2F, а ген G2D/G5 – на гены G2D и G5. Таким образом, PLA2 типа G2, по-видимому, является парафилитической группой, поскольку включает в себя и кластер PLA2 типа G5. Все остальные группы sPLA2 являются монофилитическими.

### Заключение

В нашей статье проведен анализ семейства фосфолипаз А2 человека. Мы описали структуру и функции 29 PLA2 человека, которые относятся к 12 типам: G1–8, G10, G12, G15, G16. Анализ локализации генов PLA2 в геноме человека показал, что гены фосфолипаз А2 локализованы на 12 хромосомах человека и некоторые из них образуют кластеры. Два самых крупных из них включают: первый – все гены PLA2 типа G2 (G2A, G2C–F) и ген G5, а второй – гены PLA2 типа G4 (G4B, G4D–F).

Мы также провели анализ связи между фосфолипазами А2 и заболеваниями человека по литературным данным. Было описано 229 связей «болезнь–ген» фосфолипазы А2, в результате чего выявлены связи 24 генов PLA2 со 119 заболеваниями. Оказалось, что белки фосфолипаз А2 типов G4, G2 и G7 были вовлечены в наибольшее число заболеваний по сравнению с другими типами PLA2. Тогда как с наибольшим числом типов PLA2 были связаны три группы заболеваний: новообразования, болезни системы кровообращения и болезни эндокринной системы.

При оценке толерантности/нетолерантности к мутациям генов PLA2 методом RVIS обнаружено, что большинство генов типов G4 (G4B, G4C, G4D, G4E, G4F),



**Рис. 7.** Филогенетическое дерево фосфолипаз A2: типы G1, G2, G3, G5, G10, G12.

Синим цветом даны названия типов (кластеров на дереве) фосфолипаз A2 и описано, какие таксоны представлены в каждом кластере. Красным цветом и красным квадратиком выделены белки фосфолипаз A2 человека. Возле узлов дерева через косую черту приведены два типа бутстреп-поддержки для них: ultrafast bootstrap (UFBoot)/bootstrap SH-aLRT. Текстовое описание филогенетического дерева см. в Прил. 11.

G2 (G4D, G4E, G4F), как и гены типов G3 и G7, представленными одним геном, толерантны к мутациям. Большинство генов типа G6 (G6A–C, G6F), как и типов G5 и G15, представленными одним геном, напротив, нетолерантны к мутациям. Здесь стоит особо отметить, что все типы PLA2 с преобладанием генов, толерантных к мутациям, за исключением G3, тоже связаны с наибольшим количеством заболеваний: G4 – 12 групп заболеваний, G2 – 13, G7 – 15. А все не толерантные к мутациям типы PLA2 ассоциированы с меньшим числом групп заболеваний: G6 – 9 групп заболеваний, G6 – 7, G6 – 1 группа. Это позволяет предполагать, что более высокая толерантность к мутациям определенного гена PLA2 у человека обусловлена вовлечением данного гена в большее количество заболеваний (групп заболеваний). Филогенетический анализ показал, что общее происхождение можно установить только для секреторных фосфолипаз A2 (G1, G2, G3, G5, G10, G12). Остальные типы фосфолипаз A2 (G4, G6, G7, G8, G15, G16) можно считать эволюционно независимыми.

## Список литературы / References

- Филькин С.Ю., Липкин А.В., Федоров А.Н. Суперсемейство фосфолипаз: структура, функции и применение в биотехнологии. *Успехи биол. химии*. 2020;60:369-410.
- [Filkin S.Yu., Lipkin A.V., Fedorov A.N. Phospholipase superfamily: structure, functions, and biotechnological applications. *Uspekhi Biologicheskoi Khimii = Biochemistry (Moscow)*. 2020;85(Suppl.1): S177-S195. DOI 10.1134/S0006297920140096.]
- Aloulou A., Rahier R., Arhab Y., Noiriell A., Abousalham A. Phospholipases: an overview. *Methods Mol. Biol.* 2018;1835:69-105. DOI 10.1007/978-1-4939-8672-9\_3.
- Burke J.E., Dennis D.A. Phospholipase A<sub>2</sub> biochemistry. *Cardiovasc. Drugs Ther.* 2009;23(1):49-59. DOI 10.1007/s10557-008-6132-9.
- De Maria L., Vind J., Oxenbøll K.M., Svendsen A., Patkar S. Phospholipases and their industrial applications. *Appl. Microbiol. Biotechnol.* 2007;74(2):290-300. DOI 10.1007/s00253-006-0775-x.
- Dennis E.A., Cao J., Hsu Y.-H., Magriotti V., Kokotos G. Phospholipase A<sub>2</sub> enzymes: physical structure, biological function, disease implication, chemical inhibition, and therapeutic intervention. *Chem. Rev.* 2011;111(10):6130-6185. DOI 10.1021/cr200085w.
- Giresha A.S. Secretory phospholipase A2 group IIA: a potential therapeutic target in inflammation. In: Kumar D. (Ed.) Current Research

- and Trends in Medical Science and Technology. Lucknow (Uttar Pradesh, India): Department of Ortho KGMU, 2021;1:34-85.
- Guan M., Qu L., Tan W., Chen L., Wong C.-W. Hepatocyte nuclear factor-4 alpha regulates liver triglyceride metabolism in part through secreted phospholipase A<sub>2</sub> GXIIB. *Hepatology*. 2011;53(2):458-466. DOI 10.1002/hep.24066.
- Hatfield G.W., Hung S.-P., Baldi P. Differential analysis of DNA microarray gene expression data. *Mol. Microbiol.* 2003;47(4):871-877. DOI 10.1046/j.1365-2958.2003.03298.x.
- Hirsch J.A., Nicola G., McGinty G., Liu R.W., Barr R.M., Chittle M.D., Manchikanti L. ICD-10: history and context. *Am. J. Neuroradiol.* 2016;37(4):596-599. DOI 10.3174/ajnr.A4696.
- Huang D.W., Sherman B.T., Zheng X., Yang J., Imamichi T., Stephens R., Lempicki R.A. Extracting biological meaning from large gene lists with DAVID. *Curr. Protoc. Bioinformatics*. 2009;27:13.11.1-13.11.13. DOI 10.1002/0471250953.bi1311s27.
- Huang Q., Wu Y., Qin C., He W., Wei X. Phylogenetic and structural analysis of the phospholipase A<sub>2</sub> gene family in vertebrates. *Int. J. Mol. Med.* 2015;35(3):587-596. DOI 10.3892/ijmm.2014.2047.
- Karp P.D. An ontology for biological function based on molecular interactions. *Bioinformatics*. 2000;16(3):269-285. DOI 10.1093/bioinformatics/16.3.269.
- Katoh K., Toh H. Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics*. 2010;26(15):1899-1900. DOI 10.1093/bioinformatics/btq224.
- Kudo I., Murakami M. Phospholipase A<sub>2</sub> enzymes. *Prostaglandins Other Lipid Mediat.* 2002;68-69:3-58. DOI 10.1016/s0090-6980(02)00020-5.
- Leuner K., Pantel J., Frey C., Schindowski K., Schulz K., Wegat T., Maurer K., Eckert A., Müller W.E. Enhanced apoptosis, oxidative stress and mitochondrial dysfunction in lymphocytes as potential biomarkers for Alzheimer's disease. *J. Neural. Transm. Suppl.* 2007;72:207-215. DOI 10.1007/978-3-211-73574-9\_27.
- Nguyen L.-T., Schmidt H.A., von Haeseler A., Minh B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 2015;32(1):268-274. DOI 10.1093/molbev/msu300.
- Pei J., Grishin N.V. PROMALS: towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics*. 2007;23(7):802-808. DOI 10.1093/bioinformatics/btm017.
- Petrovski S., Wang Q., Heinzen E.L., Allen A.S., Goldstein D.B. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet.* 2013;9(8):e1003709. DOI 10.1371/journal.pgen.1003709.
- Shayman J.A., Tesmer J.J.G. Lysosomal phospholipase A<sub>2</sub>. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids*. 2019;1864(6):932-940. DOI 10.1016/j.bbalip.2018.07.012.
- Shimizu T. Lipid mediators in health and disease: enzymes and receptors as therapeutic targets for the regulation of immunity and inflammation. *Annu. Rev. Pharmacol. Toxicol.* 2009;49:123-150. DOI 10.1146/annurev.pharmtox.011008.145616.
- Takahashi H., Shibuya M. The vascular endothelial growth factor (VEGF)/VEGF receptor system and its role under physiological and pathological conditions. *Clin. Sci.* 2005;109(3):227-241. DOI 10.1042/CS20040370.
- Turkmen K. Inflammation, oxidative stress, apoptosis, and autophagy in diabetes mellitus and diabetic kidney disease: the Four Horsemen of the Apocalypse. *Int. Urol. Nephrol.* 2017;49(5):837-844. DOI 10.1007/s11255-016-1488-4.

#### ORCID ID

I.I. Turnaev orcid.org/0000-0002-0448-1468  
D.A. Afonnikov orcid.org/0000-0001-9738-1409

**Благодарности.** Работа поддержана бюджетным проектом № FWNR-2022-0020. Обработка данных проводилась с использованием вычислительных ресурсов ЦКП «Биоинформатика» ИЦиГ СО РАН и Суперкомпьютерного центра Новосибирского государственного университета.

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 11.09.2022. После доработки 24.11.2022. Принята к публикации 25.11.2022.

Перевод на английский язык <https://vavilov.elpub.ru/jour>

## Промоторы генов, кодирующих $\beta$ -амилазу, альбумин и глобулин пищевых растений в сравнении с непищевыми, характеризуются более низкой аффинностью к ТАТА-связывающему белку: *in silico* анализ

О.В. Вишневецкий<sup>1</sup>, И.В. Чадаева<sup>1</sup>, Е.Б. Шарыпова<sup>1</sup>, Б.М. Хандаев<sup>1</sup>, К.А. Золотарева<sup>1</sup>, А.В. Казачек<sup>1</sup>, П.М. Пономаренко<sup>1</sup>, Н.Л. Подколотный<sup>1,2</sup>, Д.А. Рассказов<sup>1</sup>, А.Г. Богомолов<sup>1</sup>, О.А. Подколотная<sup>1</sup>, Л.К. Савинкова<sup>1</sup>, Е.В. Землянская<sup>1</sup>, М.П. Пономаренко<sup>1</sup> 

<sup>1</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>2</sup> Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук, Новосибирск, Россия

 [pon@bionet.nsc.ru](mailto:pon@bionet.nsc.ru)

**Аннотация.** Принято считать, что при доместикации пищевых растений отбор шел на урожайность, технологичность переработки в продукты питания, устойчивость к патогенам и стрессовым воздействиям окружающей среды. При этом также могли оцениваться вкусовые качества продуктов питания растительного происхождения и их ценность для здоровья. Однако неясно, проводил ли человек в прошлом наряду с положительным отбором на полезные свойства растений одновременно отбор против таких вредоносных свойств, как способность вызывать аллергические реакции. Этот вопрос становится все более актуальным по мере роста аллергии населения как вызова современной медицине. В связи с этим селекционерами уже ведутся интенсивные исследования по созданию гипоаллергенных форм пищевых растений. В этой работе рассмотрены альбумин, глобулин и  $\beta$ -амилаза мягкой пшеницы *Triticum aestivum* L. (1753), идентифицированные ранее как мишени для атак иммуноглобулинов класса Е человека. Нашей целью было найти на геномном уровне следы отрицательного отбора в прошлом против гипераллергенности трех белков (альбумин, глобулин и  $\beta$ -амилаза) при одомашнивании предковых форм современных пищевых растений. Для этого мы сфокусировали поиск на сайте связывания ТАТА-связывающего белка (ТВР) как локализованном в узком районе [-70; -20] относительно старта транскрипции, консервативном, необходимом для первичной инициации транскрипции и наиболее изученном регуляторном сигнале в геномах эукариот. Ранее нами был создан свободно доступный веб-сервис Plant\_SNP\_TATA\_Z-tester для оценки величин равновесной константы диссоциации ( $K_D$ ) комплексов ТВР с проксимальными промоторами генов растений по их последовательностям ДНК длиной 90 п.о. В настоящей работе с его помощью проанализированы 363 последовательности ДНК промоторов генов 43 видов растений. Обнаружено, что пищевые растения, в сравнении с непищевыми, характеризуются достоверно более низкой аффинностью ТВР к проксимальным промоторам их генов, гомологичных генам глобулина, альбумина и  $\beta$ -амилазы мягкой пшеницы как пищевых аллергенов ( $p < 0.01$ , Z-критерий Фишера). Это свидетельствует об отборе при доместикации пищевых растений в прошлом на снижение уровня данных аллергенных белков.

Ключевые слова: пищевые аллергены; альбумин; глобулин;  $\beta$ -амилаза; ген; промотор; мягкая пшеница *Triticum aestivum* L. (1753); растения; ТАТА-связывающий белок; ТАТА-бок; доместикация; отбор; оценки *in silico*.

**Для цитирования:** Вишневецкий О.В., Чадаева И.В., Шарыпова Е.Б., Хандаев Б.М., Золотарева К.А., Казачек А.В., Пономаренко П.М., Подколотный Н.Л., Рассказов Д.А., Богомолов А.Г., Подколотная О.А., Савинкова Л.К., Землянская Е.В., Пономаренко М.П. Промоторы генов, кодирующих  $\beta$ -амилазу, альбумин и глобулин пищевых растений в сравнении с непищевыми, характеризуются более низкой аффинностью к ТАТА-связывающему белку: *in silico* анализ. *Вавиловский журнал генетики и селекции*. 2022;26(8):798-805. DOI 10.18699/VJGB-22-96

## Promoters of genes encoding $\beta$ -amylase, albumin and globulin in food plants have weaker affinity for TATA-binding protein as compared to non-food plants: *in silico* analysis

O.V. Vishnevsky<sup>1</sup>, I.V. Chadaeva<sup>1</sup>, E.B. Sharypova<sup>1</sup>, B.M. Khandaev<sup>1</sup>, K.A. Zolotareva<sup>1</sup>, A.V. Kazachek<sup>1</sup>, P.M. Ponomarenko<sup>1</sup>, N.L. Podkolodny<sup>1,2</sup>, D.A. Rasskazov<sup>1</sup>, A.G. Bogomolov<sup>1</sup>, O.A. Podkolodnaya<sup>1</sup>, L.K. Savinkova<sup>1</sup>, E.V. Zemlyanskaya<sup>1</sup>, M.P. Ponomarenko<sup>1</sup> 

<sup>1</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>2</sup> Institute of Computational Mathematics and Mathematical Geophysics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

 [pon@bionet.nsc.ru](mailto:pon@bionet.nsc.ru)

**Abstract.** It is generally accepted that during the domestication of food plants, selection was focused on their productivity, the ease of their technological processing into food, and resistance to pathogens and environmental stressors. Besides, the palatability of plant foods and their health benefits could also be subjected to selection by humans in the past. Nonetheless, it is unclear whether in antiquity, aside from positive selection for beneficial properties of plants, humans simultaneously selected against such detrimental properties as allergenicity. This topic is becoming increasingly relevant as the allergization of the population grows, being a major challenge for modern medicine. That is why intensive research by breeders is already underway for creating hypoallergenic forms of food plants. Accordingly, in this paper, albumin, globulin, and  $\beta$ -amylase of common wheat *Triticum aestivum* L. (1753) are analyzed, which have been identified earlier as targets for attacks by human class E immunoglobulins. At the genomic level, we wanted to find signs of past negative selection against the allergenicity of these three proteins (albumin, globulin, and  $\beta$ -amylase) during the domestication of ancestral forms of modern food plants. We focused the search on the TATA-binding protein (TBP)-binding site because it is located within a narrow region (between positions -70 and -20 relative to the corresponding transcription start sites), is the most conserved, necessary for primary transcription initiation, and is the best-studied regulatory genomic signal in eukaryotes. Our previous studies presented our publicly available Web service Plant\_SNP\_TATA\_Z-tester, which makes it possible to estimate the equilibrium dissociation constant ( $K_D$ ) of TBP complexes with plant proximal promoters (as output data) using 90 bp of their DNA sequences (as input data). In this work, by means of this bioinformatics tool, 363 gene promoter DNA sequences representing 43 plant species were analyzed. It was found that compared with non-food plants, food plants are characterized by significantly weaker affinity of TBP for proximal promoters of their genes homologous to the genes of common-wheat globulin, albumin, and  $\beta$ -amylase (food allergens) ( $p < 0.01$ , Fisher's Z-test). This evidence suggests that in the past humans carried out selective breeding to reduce the expression of food plant genes encoding these allergenic proteins.

Key words: food allergen; albumin; globulin;  $\beta$ -amylase; gene; promoter; common wheat *Triticum aestivum* L. (1753); plants; TATA-binding protein; TATA box; domestication; selection; *in silico* estimate.

**For citation:** Vishnevsky O.V., Chadaeva I.V., Sharypova E.B., Khandayev B.M., Zolotareva K.A., Kazachek A.V., Ponomarenko P.M., Podkolodny N.L., Rasskazov D.A., Bogomolov A.G., Podkolodnaya O.A., Savinkova L.K., Zemlyanskaya E.V., Ponomarenko M.P. Promoters of genes encoding  $\beta$ -amylase, albumin and globulin in food plants have weaker affinity for TATA-binding protein as compared to non-food plants: *in silico* analysis. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):798-805. DOI 10.18699/VJGB-22-96

## Введение

В настоящее время проблема аллергенности продуктов питания исключительно актуальна, так как наблюдающийся стремительный рост аллергизации населения становится одним из наиболее ключевых вызовов современной медицине (Prescott et al., 2022). В связи с этим современными селекционерами ведется работа в двух направлениях: (1) создание новых гипоаллергенных форм сельскохозяйственных пищевых растений и (2) выявление новых пищевых аллергенов растений и молекулярных механизмов их действия (Hong et al., 2021; Cavazza et al., 2022).

Цель нашего исследования заключалась в поиске на молекулярно-генетическом уровне следов отрицательного отбора против аллергенов в процессе одомашнивания предковых форм современных пищевых растений. Были изучены три пищевых аллергена мягкой пшеницы *Triticum aestivum* L. (1753):  $\beta$ -амилаза, альбумин и глобулин, идентифицированные как мишени атаки иммуноглобулинов класса E человека (Wang et al., 2021).

Работа выполнялась с использованием созданного нами ранее свободно доступного веб-сервиса Plant\_SNP\_TATA\_Z-tester, который предназначен для оценки величин равновесной константы диссоциации ( $K_D$ ) комплексов TBP-1 из резуховидки Таля *Arabidopsis thaliana* (L.) Heynh., 1842 (здесь и далее – растительный TBP) с проксимальными промоторами генов растений (Rasskazov et al., 2022). С его помощью проанализированы 363 нуклеотидные последовательности проксимальных промоторов генов 43 видов растений. Обнаружено, что пищевые растения, в сравнении с непищевыми, характеризуются значимо

более низкими оценками аффинности растительного TBP к промоторам генов, гомологичных  $\beta$ -амилазе, альбумину и глобулину мягкой пшеницы как пищевым аллергенам. Это указывает на имевший место искусственный отбор, направленный на снижение уровня экспрессии генов пищевых растений, кодирующих аллергенные белки, который осуществлялся человеком в прошлом по мере одомашнивания таких растений.

## Материалы и методы

**Нуклеотидные последовательности промоторов растительных генов, проанализированных в работе.** Рассмотрены три аллергенных белка мягкой пшеницы *T. aestivum*:  $\beta$ -амилаза, альбумин и глобулин, которые ранее были экспериментально идентифицированы в качестве мишеней для атаки иммуноглобулинов класса E человека (Wang et al., 2021). Из базы данных GenBank (Benson et al., 2015) были экстрагированы нуклеотидные последовательности проксимальных промоторов длиной 90 п. о., расположенных непосредственно перед стартами транскрипции растительных генов, гомологичных генам  $\beta$ -амилазы, альбумина и глобулина мягкой пшеницы *T. aestivum*. После исключения ДНК-последовательностей промоторов с неопределенными нуклеотидами, w, s, r, y, k, m, b, d, h, v и n (согласно номенклатуре (IUPAC-IUB..., 1970)), остались 363 промоторных последовательности, относящиеся к 43 видам растений. Затем все 43 вида растений были разделены на две перекрывающиеся группы: группа I, представленная 235 проксимальными промоторами 28 видов пищевых растений, для каждого из которых удалось найти сведения об их многовековом использо-

**Таблица 1.** Характеристика 235 нуклеотидных последовательностей проксимальных промоторов генов пищевых растений – гомологов глобулина (*Glo*), альбумина (*Alb*) и  $\beta$ -амилазы (*Bmy*) мягкой пшеницы *T. aestivum*, изученных в этой работе

Виды пищевых растений		Количество промоторов		
№ п/п	Название	<i>Glo</i>	<i>Alb</i>	<i>Bmy</i>
1	Гречиха посевная <i>Fagopyrum esculentum</i> Moench, 1794	1	–	–
2	Гинкго двулопастный <i>Ginkgo biloba</i> L., 1771	1	–	–
3	Вишня Ёсино <i>Prunus yedoensis</i> var. <i>nudiflora</i> Koehne, 1912	1	–	2
4	Кукуруза <i>Zea mays</i> L., 1753	1	–	–
5	Овес посевной <i>Avena sativa</i> L., 1753	2	–	–
6	Восковница красная <i>Morella rubra</i> Siebold & Zucc.	2	2	1
7	Киноа <i>Chenopodium quinoa</i> Willd., 1798	2	–	–
8	Рис посевной <i>Oryza sativa</i> L., 1753	3	–	–
9	Дыня <i>Cucumis melo</i> L., 1753	4	2	6
10	Артишок испанский <i>Cynara cardunculus</i> L.	4	–	–
11	Дуб пробковый <i>Quercus suber</i> L.	4	–	–
12	Виноград <i>Vitis vinifera</i> L.	9	–	9
13	Кофе конголезский <i>Coffea canephora</i> Pierre ex A. Froehner, 1897	1	–	–
14	Перец стручковый <i>Capsicum annuum</i> L., 1753	26	8	27
15	Кунжут <i>Sesamum indicum</i> L.	–	1	–
16	Киви <i>Actinidia rufa</i> Franch. & Sav.	–	1	–
17	Бразильский орех <i>Bertholletia excelsa</i> Humb. & Bonpl.	–	1	–
18	Соя культурная <i>Glycine max</i> (L.) Merr., 1917	–	2	–
19	Горох посевной <i>Pisum sativum</i> L., 1753	–	4	–
20	Перилла кустарниковая <i>Perilla frutescens</i> var. <i>hirtella</i> (Nakai) Makino	–	5	–
21	Миндаль обыкновенный <i>Prunus dulcis</i> (Mill.) D.A. Webb, 1967	–	8	4
22	Мандарин уншиу <i>Citrus unshiu</i> (Tanaka ex Swingle) Marcow., 1921	–	15	–
23	Чай <i>Camellia sinensis</i> (L.) Kuntze, 1887	–	–	1
24	Ячмень обыкновенный <i>Hordeum vulgare</i> L. (1753)	–	–	2
25	Гибискус сирийский <i>Hibiscus syriacus</i> L. (1753)	–	–	2
26	Ананас настоящий <i>Ananas comosus</i> (L.) Merr., 1917	–	–	3
27	Олива европейская <i>Olea europaea</i> L., 1753	–	–	4
28	Полынь однолетняя <i>Artemisia annua</i> L.	13	35	16
Количество исследованных видов пищевых растений		15	12	12

вании людьми в качестве продуктов питания (табл. 1), и группа II, представленная 128 проксимальными промоторами непищевых растений (остальные 15 видов) (табл. 2).

**Анализ нуклеотидных последовательностей проксимальных промоторов растений.** С помощью созданного нами ранее веб-сервиса Plant\_SNP\_TATA\_Z-tester (Rasskazov et al., 2022) по нуклеотидной последовательности каждого промотора, охарактеризованного в табл. 1 и 2, рассчитывали в единицах моли на литр (М) величины  $K_D$  комплексов растительного ТВР с этим промотором.

Расчеты выполняли в соответствии с разработанной нами ранее моделью трехэтапного связывания ТВР с промотором: (i) ТВР скользит вдоль двойной спирали ДНК промотора (Coleman, Pugh, 1995)  $\leftrightarrow$  (ii) ТВР останавлива-

ется на потенциальном сайте ТВР-связывания (Berg, von Hippel, 1987; Bucher, 1990)  $\leftrightarrow$  (iii) комплекс ТВР/промотор стабилизируется за счет изгиба оси двойной спирали ДНК под прямым углом (Flatters, Lavery, 1998). Впоследствии такое трехэтапное связывание экспериментально наблюдали *in vitro* R.F. Delgadillo с коллегами (2009).

**Статистический анализ.** С использованием стандартного пакета программ Statistica (Statsoft™, США) усредняли полученные с помощью Plant\_SNP\_TATA\_Z-tester (Rasskazov et al., 2022) оценки  $K_D$  комплексов растительного ТВР к промоторам генов  $\beta$ -амилазы, альбумина и глобулина для пищевых и непищевых растений по отдельности. На этой основе оценивали статистическую значимость различия между пищевыми и непищевыми растениями согласно Z-критерию Фишера.

**Таблица 2.** Характеристика 128 нуклеотидных последовательностей проксимальных промоторов генов непищевых растений – гомологов глобулина (*Glo*), альбумина (*Alb*) и β-амилазы (*Bmy*) мягкой пшеницы *T. aestivum*, изученных в этой работе

Виды непищевых растений		Количество промоторов		
№ п/п	Название	<i>Glo</i>	<i>Alb</i>	<i>Bmy</i>
1	Маклея сердцевидная <i>Macleaya cordata</i> (Willd.) R. Br.	1	–	–
2	Стрига желтая <i>Striga asiatica</i> (L.) Kuntze	1	1	1
3	Генлиси <i>Genlisea aurea</i> A.St. Hil. (1833)	1	–	2
4	Теосинте флоридский <i>Zea luxurians</i> (Durieu & Asch.) R.M. Bird, 1978	1	–	–
5	Ярутка <i>Microthlaspi erraticum</i> (Jord.) T. Ali & Thines, 2016	2	1	2
6	Трипсакум ежевидный <i>Tripsacum dactyloides</i> (L.) L., 1759	2	–	–
7	Теосинте балассаский <i>Zea mays</i> subsp. <i>parviglumis</i> Ilitis & Doebley, 1980	3	–	–
8	Резуховидка Тяля <i>Arabidopsis thaliana</i> (L.) Heynh., 1842	4	5	3
9	Дихантелиум <i>Dichantheium oligosanthes</i> (Schult.) Gould	6	–	6
10	Водяная лилия <i>Nymphaea thermarum</i> Eb. Fisch., 1988	8	1	9
11	Василистник анемоновидный <i>Thalictrum thalictroides</i> (L.) A.J. Eames & B. Boivin	9	15	6
12	Томат чилийский <i>Solanum chilense</i> (Dunal) Reiche	15	13	6
13	Стрига египетская <i>Striga hermonthica</i> (Delile) Benth.	–	1	–
14	Вшивосемянник японский <i>Phtheirospermum japonicum</i> (Thunb.) Kanitz	–	–	1
15	Китайская роза <i>Rosa chinensis</i> Jacq., 1768	–	–	2
Количество видов непищевых растений		12	7	10

## Результаты

### Глобулин

Результаты оценки *in silico*  $K_D$  комплексов растительного ТВР с 74 проксимальными промоторами генов глобулинов из 15 видов пищевых растений в сравнении с 53 такими промоторами из 12 видов непищевых растений, полученные с использованием Plant\_SNP\_TATA\_Z-tester (Rasskazov et al., 2022), представлены в табл. 3. Согласно этой таблице, у пищевых растений оценки  $K_D$  комплексов растительного ТВР с промоторами этих генов варьировали от  $1.67 \pm 0.12$  до  $6.75 \pm 5.23$  нМ при среднем значении, равном  $2.97 \pm 0.21$  нМ, а у непищевых – от  $1.25 \pm 0.06$  до  $3.33 \pm 0.23$  нМ при среднем  $2.15 \pm 0.08$  нМ.

На рис. 1 показано сравнение среднеарифметических оценок  $K_D$  комплексов растительных ТВР к промоторам генов, кодирующих глобулины, для двух контрастных выборок (пищевых и непищевых растений) с помощью Z-критерия Фишера. Различие между сравниваемыми выборками было достоверным с  $Z = 3.59$ , с уровнем статистической значимости  $p < 0.001$ .

### Альбумин

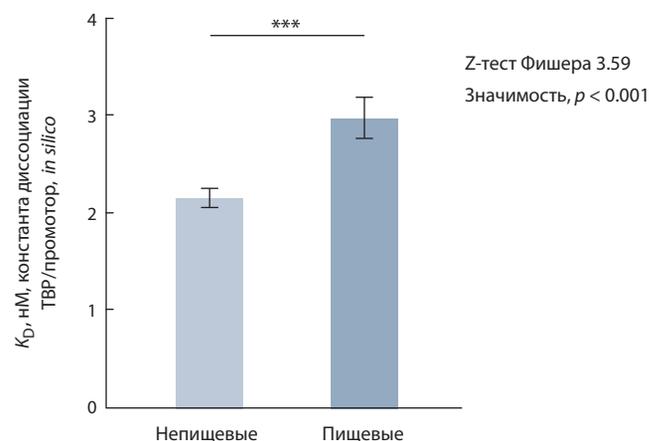
В табл. 4 приведены полученные с помощью веб-сервиса Plant\_SNP\_TATA\_Z-tester (Rasskazov et al., 2022) оценки  $K_D$  комплексов растительного ТВР с 84 промоторами генов альбуминов из 12 видов пищевых растений и с 37 промоторами из 7 видов непищевых растений. Как можно видеть из этой таблицы, оценки  $K_D$  комплексов ТВР/промотор для этих генов в случае пищевых растений наблюдались от  $1.65 \pm 0.12$  до  $4.49 \pm 1.39$  нМ (в среднем

$3.10 \pm 0.22$  нМ), тогда как у непищевых растений они были от  $1.65 \pm 0.05$  до  $2.70 \pm 0.22$  нМ (в среднем  $2.18 \pm 0.10$  нМ).

Сравнение исследуемых выборок пищевых и непищевых растений с помощью Z-критерия Фишера представлено на рис. 2: пищевые растения достоверно отличаются от непищевых ( $Z = 3.85$ ,  $p < 0.001$ ).

### β-Амилаза

В табл. 5 приведены результаты оценки  $K_D$  комплексов растительного ТВР с 77 проксимальными промоторами



**Рис. 1.** Достоверное различие исследованных выборок пищевых и непищевых растений по оценке *in silico*  $K_D$  комплексов растительного ТВР к промоторам их генов, кодирующих глобулины.

Здесь и на рис. 2: \*\*\* достоверность  $p < 0.001$  согласно Z-критерию Фишера.

**Таблица 3.** Среднеарифметические оценки величин равновесной константы диссоциации ( $K_D$ ) комплексов ТВР с проксимальными промоторами генов растительных глобулинов

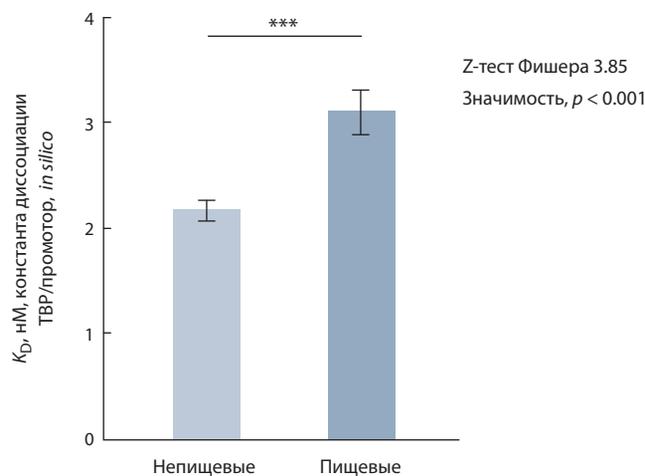
№ п/п	Вид растения	N	$K_D, M_0 \pm \Delta, \text{нМ}$
<b>Пищевые растения</b>			
1	Кофе конголезский	1	$2.17 \pm 0.13$
2	Гречиха посевная	1	$2.04 \pm 0.14$
3	Гинкго двулопастный	1	$1.67 \pm 0.12$
4	Вишня Ёсино	1	$1.76 \pm 0.12$
5	Кукуруза	1	$2.30 \pm 0.16$
6	Овес посевной	2	$2.57 \pm 0.15$
7	Восковница красная	2	$6.75 \pm 5.23$
8	Киноа	2	$2.47 \pm 0.12$
9	Рис посевной	3	$2.84 \pm 0.38$
10	Дыня	4	$2.66 \pm 0.30$
11	Артишок испанский	4	$3.22 \pm 0.37$
12	Дуб пробковый	4	$4.84 \pm 1.41$
13	Виноград	9	$2.67 \pm 0.35$
14	Полынь однолетняя	13	$2.51 \pm 0.25$
15	Перец стручковый	26	$3.01 \pm 0.37$
Итого		74	$2.97 \pm 0.21$
<b>Непищевые растения</b>			
1	Маклея сердцевидная	1	$1.25 \pm 0.06$
2	Стрига желтая	1	$3.33 \pm 0.23$
3	Генлиси	1	$2.70 \pm 0.19$
4	Теосинте флоридский	1	$1.96 \pm 0.14$
5	Ярутка	2	$2.25 \pm 0.83$
6	Трипсакум ежевидный	2	$2.19 \pm 0.07$
7	Теосинте балсасский	3	$2.12 \pm 0.11$
8	Резуховидка Таля	4	$1.90 \pm 0.13$
9	Дихантелия	6	$2.27 \pm 0.15$
10	Водяная лилия	8	$2.60 \pm 0.39$
11	Василистник анемоновидный	9	$2.01 \pm 0.16$
12	Томат чилийский	15	$1.97 \pm 0.11$
Итого		53	$2.15 \pm 0.08$

Примечание. Здесь и в табл. 4 и 5: N – количество промоторов;  $M_0$  – среднеарифметическая оценка;  $\Delta$  – стандартная ошибка.

генов  $\beta$ -амилазы из 12 видов пищевых растений, а также оценки для 38 промоторов из 10 видов непищевых растений согласно расчетам веб-сервиса Plant\_SNP\_TATA\_Z-tester (Rasskazov et al., 2022). Эта таблица характеризует пищевые растения диапазоном  $K_D$  от  $1.30 \pm 0.09$  до  $8.77 \pm 7.36$  нМ при среднеарифметической оценке  $2.85 \pm 0.21$  нМ, тогда как у непищевых растений диапазон  $K_D$  был от  $1.66 \pm 0.32$  до  $6.75 \pm 5.23$  нМ при среднем

**Таблица 4.** Среднеарифметические оценки  $K_D$  комплексов ТВР с проксимальными промоторами генов растительных альбуминов

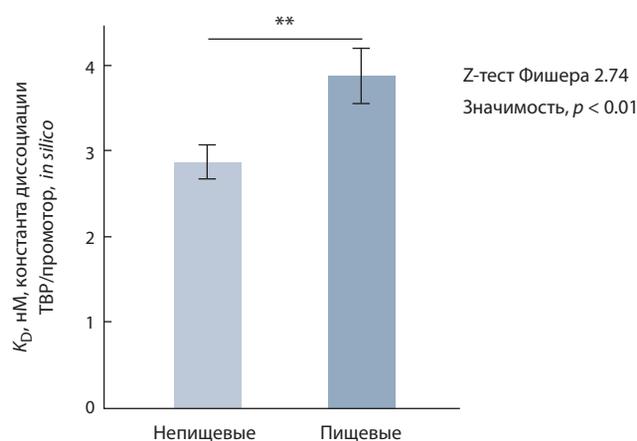
№ п/п	Вид растения	N	$K_D, M_0 \pm \Delta, \text{нМ}$
<b>Пищевые растения</b>			
1	Кунжут	1	$2.28 \pm 0.16$
2	Киви	1	$1.77 \pm 0.12$
3	Бразильский орех	1	$3.04 \pm 0.15$
4	Восковница красная	2	$1.96 \pm 0.14$
5	Дыня	2	$2.04 \pm 0.14$
6	Соя культурная	2	$1.65 \pm 0.12$
7	Горох посевной	4	$4.49 \pm 1.39$
8	Перилла кустарниковая	5	$1.98 \pm 0.40$
9	Миндаль обыкновенный	8	$3.74 \pm 0.66$
10	Перец стручковый	8	$3.00 \pm 0.59$
11	Мандарин уншиу	15	$3.51 \pm 0.61$
12	Полынь однолетняя	35	$3.07 \pm 0.35$
Итого		84	$3.10 \pm 0.22$
<b>Непищевые растения</b>			
1	Водяная лилия	1	$2.70 \pm 0.22$
2	Ярутка	1	$2.00 \pm 0.14$
3	Стрига египетская	1	$1.65 \pm 0.12$
4	Стрига желтая	1	$2.19 \pm 0.15$
5	Резуховидка Таля	5	$2.03 \pm 0.21$
6	Томат чилийский	13	$2.33 \pm 0.20$
7	Василистник анемоновидный	15	$2.11 \pm 0.16$
Итого		37	$2.18 \pm 0.10$



**Рис. 2.** Достоверное различие исследованных выборок пищевых и непищевых растений по оценке *in silico*  $K_D$  комплексов ТВР с проксимальными промоторами генов растительных ТВР к промоторам их генов, кодирующих альбумины.

**Таблица 5.** Среднеарифметические оценки  $K_D$  комплексов ТВР с проксимальными промоторами генов растительных β-амилазы

№ п/п	Вид растения	N	$K_D, M_0 \pm \Delta, \text{нМ}$
<b>Пищевые растения</b>			
1	Чай	1	$4.59 \pm 0.28$
2	Восковница красная	1	$2.21 \pm 0.13$
3	Ячмень обыкновенный	2	$1.30 \pm 0.09$
4	Гибискус сирийский	2	$3.58 \pm 1.86$
5	Вишня Ёсино	2	$3.19 \pm 1.68$
6	Ананас настоящий	3	$8.77 \pm 7.36$
7	Миндаль обыкновенный	4	$6.56 \pm 1.63$
8	Олива европейская	4	$5.24 \pm 0.93$
9	Дыня	6	$4.79 \pm 0.96$
10	Виноград	9	$3.97 \pm 0.73$
11	Полынь однолетняя	16	$4.29 \pm 0.77$
12	Перец стручковый	27	$2.59 \pm 0.23$
Итого		77	$3.89 \pm 0.32$
<b>Непищевые растения</b>			
1	Стрига желтая	1	$3.50 \pm 0.25$
2	Вшивосемянник японский	1	$3.43 \pm 0.21$
3	Генлиси	2	$1.38 \pm 0.82$
4	Ярутка	2	$1.93 \pm 0.37$
5	Китайская роза	2	$1.79 \pm 0.23$
6	Резуховидка Тая	3	$1.66 \pm 0.32$
7	Дихантелия	6	$3.26 \pm 0.43$
8	Василистник анемоновидный	6	$2.91 \pm 0.43$
9	Томат чилийский	6	$2.89 \pm 0.47$
10	Водяная лилия	9	$3.30 \pm 0.62$
Итого		38	$2.85 \pm 0.21$



**Рис. 3.** Достоверное различие исследованных выборок пищевых и непищевых растений по оценке *in silico*  $K_D$  комплексов растительного ТВР к промоторам их генов, кодирующих β-амилазы.

\*\* – Достоверность  $p < 0.01$  согласно Z-критерию Фишера.

$3.89 \pm 0.32 \text{ нМ}$ . Сравнение проанализированных выборок пищевых и непищевых растений с использованием Z-критерия Фишера, согласно которому они статистически достоверно отличаются друг от друга при  $Z = 2.74$  с уровнем статистической значимости  $p < 0.01$ , продемонстрировано на рис. 3.

### Обсуждение

Общеизвестно, что в процессе стихийной доместикации предковых форм современных пищевых растений отбор шел, прежде всего, на их хозяйственно ценные признаки, такие как «урожайность», «устойчивость к патогенам и стрессовым воздействиям окружающей среды», а также «технологичность переработки в конечные продукты питания». В процессе одомашнивания также оценивались вкусовые качества продуктов питания и их значение для здоровья.

Остается неясным, проводился ли наряду с положительным отбором на полезные свойства сельскохозяйственных растений одновременно отбор против их вредоносных свойств, к числу которых можно отнести способность приготовленных из них пищевых продуктов вызывать аллергические реакции у человека. Для ответа на этот вопрос мы сконцентрировали внимание на поиске молекулярно-генетических маркеров селекции, ассоциированных со структурно-функциональной организацией проксимальных промоторов растительных генов.

Проанализированы растительные гены – гомологи трех генов мягкой пшеницы *T. aestivum*, кодирующих пищевые аллергены – β-амилазу, альбумин и глобулин, идентифицированные как мишени для атаки иммуноглобулинов класса E человека (Wang et al., 2021). Рассмотрены 363 гена-гомолога, относящихся к 28 и 15 видам пищевых и непищевых растений соответственно. Для каждого гена-гомолога с использованием веб-сервиса Plant\_SNP\_TATA\_Z-tester (Rasskazov et al., 2022) была рассчитана величина  $K_D$  комплекса растительного ТВР с проксимальным промотором этого гена.

Интерес к белку ТВР и сайту его связывания в проксимальном промоторе (каноническая форма: ТАТА-бокс) обусловлен тем, что они играют ключевую роль в инициации транскрипции генов эукариот. Экспериментально установлено (Coleman, Pugh, 1995), что ТВР скользит вдоль двойной спирали ДНК благодаря неспецифической аффинности между ними,  $K_D \sim 10^{-5} \text{ М}$  (Hahn et al., 1989). Затем ТВР останавливается на сайте ТВР-связывания в силу их взаимного молекулярного распознавания (Berg, von Hippel, 1987; Bucher, 1990) из-за большей специфической аффинности ТВР к этому сайту,  $K_D \sim 10^{-9} \text{ М}$  (Hahn et al., 1989). Далее под действием ТВР происходят плавление двойной спирали ДНК в сайте ТВР-связывания и излом ее оси под прямым углом, что стабилизирует комплекс ТВР/промотор (Flatters, Lavery, 1998). Сформировавшийся комплекс ТВР/промотор принято считать обязательным «якорем» на ДНК, который необходим для связывания РНК полимеразы II (Muller et al., 2001; Martianov et al., 2002; Choukrallah et al., 2012; Rhee, Pugh, 2012) в качестве ключевого шага сборки преиницирующего комплекса транскрипции (Auble, 2009), ответственного за базальную транскрипцию (Fire et al., 1984). Вследствие ключевой

значимости ТАТА-боксов мутации, локализованные в проксимальных промоторах, оказывают выраженное влияние на уровень экспрессии генов (Савинкова и др., 2009).

Молекулярный механизм связывания ТВР с промоторами генов эукариот за три последовательных шага впервые был предложен в работе (Пономаренко и др., 2008) и позднее подтвержден экспериментально (Delgadillo et al., 2009). На основе этого механизма разработана биоинформатическая модель для расчета изменения  $K_D$  комплексов ТВР с проксимальными промоторами генов эукариот в случае полиморфизма сайтов ТВР-связывания в этих промоторах (Пономаренко и др., 2009). Результаты расчетов с использованием этой модели были подтверждены в независимых экспериментах *ex vivo* на культурах клеток, трансфицированных плазмидой pGL4.10 (Promega, США), несущей нормальный или мутантный вариант исследуемого промотора перед репортерным геном люциферазы (Ponomarenko et al., 2017), а также *in vitro* в реальном времени (Arkova et al., 2017) на спектрометре остановленной струи SX.20 (Applied Photophysics, Великобритания), в равновесных (Savinkova et al., 2013) и неравновесных (Drachkova et al., 2014) условиях гелретадации. В итоге столь разносторонней верификации этой биоинформатической модели на ее основе был создан веб-сервис Plant SNP\_TATA\_Z-tester (Rasskazov et al., 2022), использованный в данной работе для оценки  $K_D$  комплексов растительного ТВР с проксимальными промоторами генов пищевых и непищевых растений.

Осуществленный нами анализ показал, что пищевые растения, в сравнении с непищевыми, характеризуются значимо более низкими оценками аффинности ТВР к промоторам генов, гомологичных  $\beta$ -амилазе, альбумину и глобулину мягкой пшеницы как пищевым аллергенам ( $p < 0.01$ , Z-критерий Фишера). Рассматривая полученные результаты, примем во внимание тот экспериментально установленный факт, что уровень экспрессии генов эукариот повышается при увеличении аффинности ТВР к промоторам данных генов (Mogno et al., 2010). Это позволяет нам интерпретировать выявленное у пищевых растений, в сравнении с непищевыми, снижение аффинности ТВР к промоторам генов, гомологичных генам  $\beta$ -амилазы, альбумина и глобулина мягкой пшеницы как пищевых аллергенов. Это может свидетельствовать о проведенном человеком в прошлом отборе пищевых растений на пониженное содержание этих аллергенных белков.

## Заключение

В настоящей работе у 28 пищевых и 15 непищевых видов растений впервые единообразно проанализированы *in silico* 363 последовательности ДНК длиной 90 п. о. проксимальных промоторов генов, гомологичных генам мягкой пшеницы, кодирующим пищевые аллергены: глобулин, альбумин и  $\beta$ -амилазу (Wang et al., 2021). В результате обнаружены пониженные оценки *in silico* аффинности ТВР к промоторам исследуемых генов пищевых растений в сравнении с непищевыми. Это может свидетельствовать в пользу имевшего место искусственного отбора пищевых растений, направленного на пониженный уровень экспрессии генов, кодирующих аллергенные белки, который осуществлялся человеком в прошлом.

## Список литературы / References

- Пономаренко П.М., Пономаренко М.П., Драчкова И.А., Лысова М.В., Аршинова Т.В., Савинкова Л.К., Колчанов Н.А. Прогноз изменения аффинности ТАТА-связывающего белка к ТАТА-боксам в результате полиморфизмов ТАТА-боксов промоторов генов человека. *Молекуляр. биология*. 2009;43(3):512-520.  
[Ponomarenko P.M., Ponomarenko M.P., Drachkova I.A., Lysova M.V., Arshinova T.V., Savinkova L.K., Kolchanov N.A. Prediction of the affinity of the TATA-binding protein to TATA boxes with single nucleotide polymorphisms. *Mol. Biol. (Moscow)*. 2009; 43(3):472-479. DOI 10.1134/S0026893309030157.]
- Пономаренко П.М., Савинкова Л.К., Драчкова И.А., Лысова М.В., Аршинова Т.В., Пономаренко М.П., Колчанов Н.А. Пошаговая модель связывания ТВР/ТАТА-бокс позволяет предсказать наследственное заболевание человека по точечному полиморфизму. *Докл. Академии наук*. 2008;419(6):828-832.  
[Ponomarenko P., Savinkova L., Drachkova I., Lysova M., Arshinova T., Ponomarenko M., Kolchanov N. A step-by-step model of TBP/TATA box binding allows predicting human hereditary diseases by single nucleotide polymorphism. *Dokl. Biochem. Biophys.* 2008; 419:88-92. DOI 10.1134/S1607672908020117.]
- Савинкова Л.К., Пономаренко М.П., Пономаренко П.М., Драчкова И.А., Лысова М.В., Аршинова Т.В., Колчанов Н.А. Полиморфизмы ТАТА-боксов промоторов генов человека и ассоциированные с ними наследственные патологии. *Биохимия*. 2009; 74(2):149-163.  
[Savinkova L.K., Ponomarenko M.P., Ponomarenko P.M., Drachkova I.A., Lysova M.V., Arshinova T.V., Kolchanov N.A. TATA box polymorphisms in human gene promoters and associated hereditary pathologies. *Biochemistry (Moscow)*. 2009;74(2):117-129. DOI 10.1134/s0006297909020011.]
- Arkova O., Kuznetsov N., Fedorova O., Savinkova L. A real-time study of the interaction of TBP with a TATA box-containing duplex identical to an ancestral or minor allele of human gene *LEP* or *TPI*. *J. Biomol. Struct. Dyn.* 2017;35(14):3070-3081. DOI 10.1080/07391102.2016.1241190.
- Auble D.T. The dynamic personality of TATA-binding protein. *Trends Biochem. Sci.* 2009;34(2):49-52. DOI 10.1016/j.tibs.2008.10.008.
- Benson D.A., Clark K., Karsch-Mizrachi I., Lipman D.J., Ostell J., Sayers E.W. GenBank. *Nucleic Acids Res.* 2015;43(Database issue):D30-D35. DOI 10.1093/nar/gku1216.
- Berg O.G., von Hippel P.H. Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. *J. Mol. Biol.* 1987;193(4):723-750. DOI 10.1016/0022-2836(87)90354-8.
- Bucher P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* 1990;212(4):563-578. DOI 10.1016/0022-2836(90)90223-9.
- Cavazza A., Mattarozzi M., Franzoni A., Careri M. A spotlight on analytical prospects in food allergens: From emerging allergens and novel foods to bioplastics and plant-based sustainable food contact materials. *Food Chem.* 2022;388:132951. DOI 10.1016/j.foodchem.2022.132951.
- Choukrallah M.A., Kobi D., Martianov I., Pijnappel W.W., Mischerikow N., Ye T., Heck A.J., Timmers H.T., Davidson I. Interconversion between active and inactive TATA-binding protein transcription complexes in the mouse genome. *Nucleic Acids Res.* 2012;40(4):1446-1459. DOI 10.1093/nar/gkr802.
- Coleman R.A., Pugh B.F. Evidence for functional binding and stable sliding of the TATA binding protein on nonspecific DNA. *J. Biol. Chem.* 1995;270(23):13850-13859. DOI 10.1074/jbc.270.23.13850.
- Delgadillo R.F., Whittington J.E., Parkhurst L.K., Parkhurst L.J. The TATA-binding protein core domain in solution variably bends TATA sequences via a three-step binding mechanism. *Biochemistry*. 2009; 48(8):1801-1809. DOI 10.1021/bi8018724.
- Drachkova I., Savinkova L., Arshinova T., Ponomarenko M., Peltek S., Kolchanov N. The mechanism by which TATA-box polymorphisms

- associated with human hereditary diseases influence interactions with the TATA-binding protein. *Hum. Mutat.* 2014;35(5):601-608. DOI 10.1002/humu.22535.
- Fire A., Samuels M., Sharp P.A. Interactions between RNA polymerase II, factors, and template leading to accurate transcription. *J. Biol. Chem.* 1984;259(4):2509-2516. DOI 10.1016/S0021-9258(17)43382-5.
- Flatters D., Lavery R. Sequence-dependent dynamics of TATA-Box binding sites. *Biophys. J.* 1998;75(1):372-381. DOI 10.1016/S0006-3495(98)77521-6.
- Hahn S., Buratowski S., Sharp P.A., Guarente L. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. *Proc. Natl. Acad. Sci. USA.* 1989; 86(15):5718-5722. DOI 10.1073/pnas.86.15.5718.
- Hong L., Pan M., Xie X., Liu K., Yang J., Wang S., Wang S. Aptamer-based fluorescent biosensor for the rapid and sensitive detection of allergens in food matrices. *Foods.* 2021;10(11):2598. DOI 10.3390/foods10112598.
- IUPAC-IUB Commission on Biochemical Nomenclature (CBN). Abbreviations and symbols for nucleic acids, polynucleotides and their constituents. Recommendations 1970. *Biochem. J.* 1970;120(3): 449-454. DOI 10.1042/bj1200449.
- Martianov I., Viville S., Davidson I. RNA polymerase II transcription in murine cells lacking the TATA binding protein. *Science.* 2002; 298(5595):1036-1039. DOI 10.1126/science.1076327.
- Mogno I., Vallania F., Mitra R.D., Cohen B.A. TATA is a modular component of synthetic promoters. *Genome Res.* 2010;20(10):1391-1397. DOI 10.1101/gr.106732.110.
- Muller F., Lakatos L., Dantonel J., Strahle U., Tora L. TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish. *Curr. Biol.* 2001;11(4):282-287. DOI 10.1016/S0960-9822(01)00076-8.
- Ponomarenko P., Chadaeva I., Rasskazov D.A., Sharypova E., Kashina E.V., Drachkova I., Zhechev D., Ponomarenko M.P., Savinkova L.K., Kolchanov N. Candidate SNP markers of familial and sporadic Alzheimer's diseases are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *Front. Aging Neurosci.* 2017;9:231. DOI 10.3389/fnagi.2017.00231.
- Prescott S.L., Logan A.C., Bristow J., Rozzi R., Moodie R., Redvers N., Haahela T., Warber S., Poland B., Hancock T., Berman B. Exiting the anthropocene: achieving personal and planetary health in the 21st century. *Allergy.* 2022;77(12):3498-3512. DOI 10.1111/all.15419.
- Rasskazov D., Chadaeva I., Sharypova E., Zolotareva K., Khandaev B., Ponomarenko P., Podkolodny N., TverdokhleB N., Vishnevsky O., Bogomolov A., Podkolodnaya O., Savinkova L., Zemlyanskaya E., Golubyatnikov V., Kolchanov N., Ponomarenko M. Plant SNP\_TATA\_Z-tester: a Web service that unequivocally estimates the impact of proximal promoter mutations on plant gene expression. *Int. J. Mol. Sci.* 2022;23(15):8684. DOI 10.3390/ijms23158684.
- Rhee H., Pugh B. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature.* 2012;483(7389):295-301. DOI 10.1038/nature10799.
- Savinkova L., Drachkova I., Arshinova T., Ponomarenko P., Ponomarenko M., Kolchanov N. An experimental verification of the predicted effects of promoter TATA-box polymorphisms associated with human diseases on interactions between the TATA boxes and TATA-binding protein. *PLoS One.* 2013;8(2):e54626. DOI 10.1371/journal.pone.0054626.
- Wang Y., Weng J., Zhu C., Ai R., Zhou J., Wang C., Chen Q., Fu L. Allergenicity assessment and allergen profile analysis of different Chinese wheat cultivars. *World Allergy Organ. J.* 2021;14(7):100559. DOI 10.1016/j.waojou.2021.100559.

#### ORCID ID

O.V. Vishnevsky orcid.org/0000-0002-0347-2252  
I.V. Chadaeva orcid.org/0000-0002-2724-5441  
E.B. Sharypova orcid.org/0000-0002-1467-9312  
P.M. Ponomarenko orcid.org/0000-0003-2715-9612  
N.L. Podkolodny orcid.org/0000-0001-9132-7997

D.A. Rasskazov orcid.org/0000-0003-4795-0954  
A.G. Bogomolov orcid.org/0000-0003-4359-6089  
O.A. Podkolodnaya orcid.org/0000-0003-3247-0114  
L.K. Savinkova orcid.org/0000-0003-4543-4104  
E.V. Zemlyanskaya orcid.org/0000-0002-4304-1190  
M.P. Ponomarenko orcid.org/0000-0003-1663-318X

**Благодарности.** Работа выполнена при поддержке Российского научного фонда, грант № 20-14-00140. Авторы выражают благодарность Центру коллективного пользования «Биоинформатика» за использование вычислительных ресурсов в рамках бюджетного проекта Правительства РФ FWNr-2022-0020 и Федеральной научно-технической программы развития генетических технологий России.

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 16.09.2022. После доработки 29.11.2022. Принята к публикации 30.11.2022.

# FastContext: A tool for identification of adapters and other sequence patterns in next generation sequencing (NGS) data

E. Viesná<sup>1,2</sup>, V. Fishman<sup>1,2</sup> 

<sup>1</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

 minja@bionet.nsc.ru

**Abstract.** The development of next generation sequencing (NGS) methods has created the need for detailed analysis and control of each protocol step. NGS library preparation protocols may include steps with incorporation of various service sequences, such as sequencing adapters, primers, sample-, cell-, and molecule-specific barcodes. Despite a fairly high level of current knowledge, during the protocol development process researchers often have to deal with various kinds of unexpected experiment outcomes, which result either from lack of information, lack of knowledge, or defects in reagent manufacturing. Detection and analysis of service sequences, their distribution and linkage may provide important information for protocol optimization. Here we introduce FastContext, a tool designed to analyze NGS read structure, based on sequence features found in reads, and their relative position in the read. The algorithm is able to create human readable read structures with user-specified patterns, to calculate counts and percentage of every read structure. Despite the simplicity of the algorithm, FastContext may be useful in read structure analysis and, as a result, can help better understand molecular processes that take place at different stages of NGS library preparation. The project is open-source software, distributed under GNU GPL v3, entirely written in the programming language Python, and based on well-maintained packages and commonly used data formats. Thus, it is cross-platform, may be patched or upgraded by the user if necessary. The FastContext package is available at the Python Package Index (<https://pypi.org/project/FastContext>), the source code is available at GitHub (<https://github.com/regnveig/FastContext>).

Key words: next generation sequencing; NGS; adapters; patterns search; read analysis.

**For citation:** Viesná E., Fishman V. FastContext: A tool for identification of adapters and other sequence patterns in next generation sequencing (NGS) data. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):806-809. DOI 10.18699/VJGB-22-97

## FastContext: инструмент для контекстного анализа последовательностей в данных секвенирования нового поколения (NGS)

Э. Весна<sup>1,2</sup>, В.С. Фишман<sup>1,2</sup> 

<sup>1</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>2</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

 minja@bionet.nsc.ru

**Аннотация.** Бурное развитие методов секвенирования нового поколения (next generation sequencing, NGS) породило потребность в детальном анализе и контроле качества на каждом этапе протокола приготовления геномных библиотек. Протоколы могут включать в себя этапы с внедрением различного рода служебных последовательностей, таких как адаптеры, праймеры, а также баркоды, специфичные для каждого образца, клетки или молекулы ДНК. Несмотря на достаточно высокий уровень современных знаний в молекулярной биологии, в процессе разработки протоколов NGS исследователи часто сталкиваются с неожиданными экспериментальными данными, которые могут быть результатом недостатка информации о молекулярных процессах, сопровождающих приготовление геномных библиотек, или, в отдельных случаях, дефектом производства реактивов. Обнаружение и анализ распределения служебных последовательностей в полученных молекулах ДНК могут быть важным источником информации, необходимой для оптимизации протокола приготовления геномных библиотек. В настоящей статье представлена утилита FastContext, с помощью которой возможен анализ структуры прочтений с точки зрения присутствия определенных последовательностей и их взаимного расположения в прочтении. Алгоритм принимает на вход необработанные данные секвенирования в формате FastQ, а затем генерирует удобные для интерпретации представления структуры прочтений на основе заданных пользователем паттернов, высчитывает количество подобных структур и их долю от общего числа прочтений. Несмотря на простоту алгоритма, FastContext может быть полезен при анализе структуры прочтений, он помогает лучше

понять молекулярные процессы, происходящие на разных стадиях приготовления геномных библиотек и, как следствие, открывает возможности для усовершенствования протокола. FastContext – это проект с открытым исходным кодом, распространяемый под свободной лицензией GNU General Public License v3, полностью написанный на языке программирования Python и основанный на широко используемых программных пакетах и форматах данных. Таким образом, он может быть легко использован под любой операционной системой, исправлен и дополнен при необходимости. FastContext доступен в виде пакета в Python Package Index (<https://pypi.org/project/FastContext>), исходный код хранится на GitHub (<https://github.com/regnevig/FastContext>).

Ключевые слова: секвенирование нового поколения; NGS; адаптеры; поиск паттернов; анализ прочтений.

## Introduction

Since the advent of next generation sequencing (NGS) methods 20 years ago, those methods have been actively evolving and are currently applied to various areas of biology. Due to the increasing capacity of sequencers, it is now possible to obtain billions of short molecule sequences in a single NGS run. In order to utilize such a high throughput of modern sequencers, there is a practice of sample pooling. This method requires incorporation of sample-specific service sequences (barcodes), which allow to distinguish individual samples in raw sequencing data.

Other types of service sequences could be incorporated into the target molecules, such as sequencing adapters and primers, biotin-labeled oligonucleotides for target molecules enrichment (Gridina et al., 2021), molecule- and cell-specific barcodes, which are designed to identify a molecule (Smirnov et al., 2020) and/or a cell of origin (Aldridge, Teichmann, 2020).

There are many strategies in molecular genetics that are used for service sequences incorporation: direct ligation of DNA or RNA molecules, template-switching activity of reverse transcriptases, and incorporation of synthetic DNA transposons. During the whole process of new NGS methods development it is crucial to control each protocol step. In light of that, detection and analysis of service sequences distribution may provide important information for protocol optimization.

Here we introduce the FastContext tool, which is designed to analyze and compute statistics on NGS read structures. FastContext allows to search for user-specified sequences in NGS reads, gather data on their linkage, frequency of occurrence, and present statistics in a user-friendly manner.

## Materials and methods

The script is completely written in the programming language Python (version 3.8). It is packaged as a part of the Python Package Index (<https://pypi.org/project/FastContext>) and can be installed via pip. Therefore, it works out of the box on every operating system.

We used the following Python libraries:

1. bioPython, version 1.79 (Cock et al., 2009): FastQ files parsing and sequences manipulation;
2. python-Levenshtein<sup>1</sup>, version 0.12.2: calculating sequences Levenshtein distance;
3. pandas, version 1.2.5 (The Pandas Development Team, 2020): tables creation;
4. tqdm, version 4.61.2 (Costa-Luis et al., 2022): visualization.

All libraries listed above, except python-Levenshtein, are widely used and well maintained.

FastContext supports multi-processing.

<sup>1</sup> Available at: <https://github.com/ztane/python-Levenshtein>.

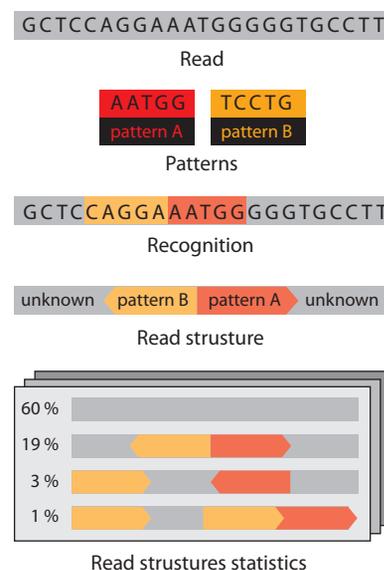


Fig. 1. FastContext algorithm scheme.

Two different example patterns colored as red and yellow.

## Results

We developed an algorithm which parses raw sequencing dataset, searches each read or read pair for user specified patterns, and then generates a human-readable representation of the search results, which we call “read structure”. Algorithm scheme is represented in the Fig. 1.

### Input

Input files are provided in FastQ<sup>2</sup> format. The user can provide one (in the single-end mode) or two (in the paired-end mode) FastQ files. Files may be uncompressed or compressed with gzip or bz2 algorithms.

### Output

Output results are provided as an HTML page (further: “summary file”), containing run options and tables with read structures, their counts, and percentages (Fig. 2). The sequence strand (forward F, or reverse R) is displayed after a colon (e.g., {**oligb:F**}).

The user can manually set minimal rate value (rate floor) to be displayed. Also, the user can save the read structure for each read or read pair, with the read name, the sequence, and Phred qualities, as a gzip-compressed JavaScript Object Notation (JSON)<sup>3</sup> object (further: “detailed statistics file”).

<sup>2</sup> Full specification of FastQ format is available at <http://maq.sourceforge.net/fastq.shtml>.

<sup>3</sup> Full specification of JSON format could be found at JSON official website: <https://www.json.org>.

Count	Percentage	Read structure
5,197	48.80	{unknown}
3,297	30.96	{unknown}--{oligme:F}--{oligb:F}--{701:F}--{unknown}
114	1.07	{unknown}--{oligb:F}--{701:F}--{unknown}
71	0.66	{unknown}--{oligme:F}--{unknown}
69	0.64	{unknown}--{oligme:F}--{unknown}--{701:F}--{unknown}
60	0.56	{unknown}--{oligme:F}--{oligb:F}--{701:F}--{kmer:14bp}

**Fig. 2.** Example of statistics table.

Every fragment of read structure, except palindromic or unrecognized sequences, has a strand suffix. Short unrecognized sequences (K-mers) have a length suffix.

## Patterns

Pattern names and sequences are provided as a plain JSON object, e. g.:

```
{"foo": "CTGTCTCTTATACAC", "bar": "CCGAAAACACG", "baz": "TCGTCGGG"}.
```

It should be noted that pattern sequences are searched in the order they are provided by the user, forward strand (the sequence itself) first, reverse strand (a reverse complement of the sequence) after. Therefore, the patterns order matters in search and should be carefully considered before running the program. FastContext expects patterns to be sorted from long to short, which is the best option for overlapping or nested sequences, and otherwise gives you a warning.

## K-mers

FastContext performs the search based on full match, and a pattern sequence with one single sequencing error will be skipped as an unrecognized sequence (alias {unknown}). This is especially important for long patterns, which are under-represented due to higher cumulative frequency of sequencing errors. In addition, oligonucleotide synthesis errors and some enzymatic steps of NGS library preparation, such as A-tailing, may produce molecules one base pair shorter or longer than expected. In order to simplify identification of such extended or truncated sequences, we have implemented the ability to mark short unrecognized sequences (K-mers) of certain length (e. g., {kmer:14bp}). If a K-mer identified in the read is one base longer or shorter than a pattern sequence, we can suppose this K-mer is the pattern sequence, and test the hypothesis in a more detailed analysis of reads.

## Levenshtein distances analysis

Additional features implemented to account for sequencing errors include analysis of Levenshtein distances between different pattern sequences (pattern analysis), and between pattern sequences and read sequence. Pattern analysis is shown in the summary file, data on every single read can be found in the detailed statistics file.

Analysis of distances between pattern sequences can prevent pattern match or nesting, when sequences are confused with each other because of a few sequencing errors. Also, FastContext warns the user about palindromes and sequences that can become palindromic because of sequencing errors. This kind of sequence may affect statistics of forward-reverse orientation.

Analysis of distances between pattern and read sequences can show similarity of an unrecognized sequence and a pattern

sequence, so the user could suggest the real read structure even if FastContext fails to do that. All these data may be found in a detailed statistics file, with Levenshtein read analysis enabled (disabled by default).

## System requirements and performance

By design, FastContext stores FASTQ reads in random access memory (RAM), therefore, the only system limitation is the RAM size. Tests we have performed show that 8 Gb RAM is enough for processing 10,000 reads, which is a high enough sample size for practical application of the tool.

There are two stages that determine the time taken for completing a task. Reading data from a physical storage (HDD, SSD, etc.) depends on the storage characteristics. Read analysis is parallelized and depends on the core number. We estimated FastContext performance characteristics on the laboratory computing server with 16 cores and 50 Gb RAM. The dependence of processing speed on process count matches the expected values. 10,000 of paired-end reads are processing for 2 seconds with 4 cores used, saving JSON increases that time to 6 seconds. With Levenshtein statistics, the same data are processing for 11 seconds, and 80 seconds are required to save JSON.

## Code access

FastContext source code is available at GitHub (<https://github.com/regnveig/FastContext>) and is distributed under GNU General Public License v3.

## Discussion

Despite the simplicity of the algorithm, FastContext may be useful in read structure analysis. It has an appealing combination of cutadapt (Martin, 2011) and FastQC (Andrews, 2010) features.

Recently, A. Bravo et al. (2021) presented a tool named 2FAST2Q, which has features similar to FastContext, including extracting and counting feature occurrences in FastQ files. Unlike FastContext, 2FAST2Q can search for frequent unknown sequences (so called extract and count mode), can handle sequence mismatches, takes into account base Phred qualities, and therefore provides more accurate statistics on feature counts. The qualitative difference of FastContext is that the tool can collect statistics on relative position of features in the read and features linkage.

There remains the problem of sequencing errors. The possibility of errors is directly dependent on sequence length. FastContext performs the search based on full match, there-

fore, under equal conditions, pattern sequences of greater length have a lower chance to be found, which may impact resulting statistics.

Similarity based on Levenshtein distance is a crude approximation to probability of presence of a particular sequence. It fails to take account of *in vitro* processes during library preparation and sequencing. This problem may be solved in future versions. As for now, the user can find Phred quality scores for each read in a detailed statistics file, and estimate analysis quality manually.

Another possible feature that can be discussed is wildcards (symbols which denote more than one canonical nucleobase). This feature may be implemented in future versions.

## Conclusion

From all of the above, we can conclude that FastContext is effective as a tool for NGS data analysis, and could be a very useful source of information in the development of new molecular biology methods.

## References

- Aldridge S., Teichmann S. Single cell transcriptomics comes of age. *Nat. Commun.* 2020;11(1):4307. DOI 10.1038/s41467-020-18158-5.
- Andrews S. FastQC: A quality control tool for high throughput sequence data. 2010. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Bravo A., Typas A., Veening J. 2FASTQ: A general-purpose sequence search and counting program for FASTQ files [preprint]. *BioRxiv*. 2021. DOI 10.1101/2021.12.17.473121.
- Cock P., Antao T., Chang J., Chapman B., Cox C., Dalke A., Friedberg I., Hamelryck T., Kauff F., Wilczynski B., de Hoon M. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 2009;25(11):1422-1423. DOI 10.1093/bioinformatics/btp163.
- Costa-Luis C., Larroque S., Altendorf K., Mary H., Korobov M., Yorav-Raphael N., Ivanov I., Bargull M., Rodrigues N., Chen G., Newey C., Zugnoni M., Pagel M., Dektyarev M., Rothberg A., Lee A., Panteleit D., Dill F., Kemenade H., McCracken J., Nordlund M., Nechaev N., Desh O. tqdm: A fast, Extensible Progress Bar for Python and CLI. *Zenodo*. 2022. DOI 10.5281/zenodo.595120.
- Gridina M., Mozheiko E., Valeev E., Nazarenko L., Lopatkina M., Markova Z., Yablonskaya M., Voinova V., Shilova N., Lebedev I., Fishman V. A cookbook for DNase Hi-C. *Epigenetics Chromatin*. 2021; 14(1):15. DOI 10.1186/s13072-021-00389-5.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17(1):10-12. DOI 10.14806/ej.17.1.200.
- Smirnov A., Fishman V., Yunusova A., Korablev A., Serova I., Skryabin B., Rozhdestvensky T., Battulin N. DNA barcoding reveals that injected transgenes are predominantly processed by homologous recombination in mouse zygote. *Nucleic Acids Res.* 2020;48(2):719-735. DOI 10.1093/nar/gkz1085.
- The Pandas Development Team. pandas-dev/pandas: Pandas. *Zenodo*. 2020. DOI 10.5281/zenodo.3509134.

---

## ORCID ID

E. Viesná [orcid.org/0000-0003-3480-3963](https://orcid.org/0000-0003-3480-3963)  
V. Fishman [orcid.org/0000-0002-5573-3100](https://orcid.org/0000-0002-5573-3100)

**Acknowledgements.** This work was supported by Russian Science Foundation, grant No. 22-14-00247. High-throughput computations required for FastContext testing were performed using the Collective usage center of the Institute of Cytology and Genetics SB RAS, 121031800061-7 (Mechanisms of genetic control of development, physiological processes and behavior in animals).

**Conflict of interest.** The authors declare no conflict of interest.

Received July 18, 2022. Revised September 2, 2022. Accepted September 7, 2022.

# Unique or not unique? Comparative genetic analysis of bacterial O-antigens from the Oxalobacteraceae family

S.D. Afonnikova<sup>1, 2</sup> , A.S. Komissarov<sup>3</sup>, P.D. Kuchur<sup>3</sup>

<sup>1</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup> ITMO University, SCAMT Institute, St. Petersburg, Russia

 svetaafonnikova@gmail.com

**Abstract.** Many plants and animals have symbiotic relationships with microorganisms, including bacteria. The interactions between bacteria and their hosts result in different outcomes for the host organism. The outcome can be neutral, harmful or have beneficial effects for participants. Remarkably, these relationships are not static, as they change throughout an organism's lifetime and on an evolutionary scale. One of the structures responsible for relationships in bacteria is O-antigen. Depending on the characteristics of its components, the bacteria can avoid the host's immune response or establish a mutualistic relationship with it. O-antigen is a key component in Gram-negative bacteria's outer membrane. This component facilitates interaction between the bacteria and host immune system or phages. The variability of the physical structure is caused by the genomic variability of genes encoding O-antigen synthesis components. The genes and pathways of O-polysaccharide (OPS) synthesis were intensively investigated mostly for Enterobacteriaceae species. Considering high genetic and molecular diversity of this structure even between strains, these findings may not have caught the entire variety possibly presented in non-model species. The current study presents a comparative analysis of genes associated with O-antigen synthesis in bacteria of the Oxalobacteraceae family. In contrast to existing studies based on PCR methods, we use a bioinformatics approach and compare O-antigens at the level of clusters rather than individual genes. We found that the O-antigen genes of these bacteria are represented by several clusters located at a distance from each other. The greatest similarity of the clusters is observed within individual bacterial genera, which is explained by the high variability of O-antigens. The study describes similarities of OPS genes inherent to the family as a whole and also considers individual unique cases of O-antigen genetic variability inherent to individual bacteria.

**Key words:** O-antigen gene clusters; lipopolysaccharide genes; comparative analysis; O-antigen; Oxalobacteraceae; *Massilia*; *Collimonas*; *Janthinobacterium*; saccharide gene cluster.

**For citation:** Afonnikova S.D., Komissarov A.S., Kuchur P.D. Unique or not unique? Comparative genetic analysis of bacterial O-antigens from the Oxalobacteraceae family. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):810-818. DOI 10.18699/VJGB-22-98

## Сравнительный генетический анализ O-антигенов бактерий семейства Oxalobacteraceae: уникальность или тривиальность?

С.Д. Афонникова<sup>1, 2</sup> , А.С. Комиссаров<sup>3</sup>, П.Д. Кучур<sup>3</sup>

<sup>1</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>2</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

<sup>3</sup> Национальный исследовательский университет ИТМО, Институт SCAMT, Санкт-Петербург, Россия

 svetaafonnikova@gmail.com

**Аннотация.** Многие растения и животные способны устанавливать симбиотические взаимоотношения с микроорганизмами, в том числе с бактериями. Специфика этих взаимодействий может приводить к разным последствиям для организма-хозяина. Взаимоотношения могут быть нейтральными, негативными либо выгодными для одной или обеих сторон. Примечательно, что взаимоотношения бактерия-хозяин не являются статичными: они могут изменяться в течение жизни организмов и в ходе их эволюции. Одной из структур, определяющих направление изменчивости, является O-антиген. В зависимости от особенностей его компонентов бактерия может избегать иммунного ответа со стороны организма-хозяина, становясь патогеном, либо устанавливать с хозяином мутуалистические отношения. O-антиген – это ключевой компонент наружной мембраны грамотрицательных бактерий. Этот компонент обеспечивает взаимодействие между бактериями и иммунной системой хозяина или фагами. Варибельность структуры O-антигенов тесно связана с изменчивостью генов, кодирующих компоненты его синтеза. Гены и пути синтеза O-антигенов наиболее детально изучены у бактерий из семейства Enterobacteriaceae. С учетом высокого генетического и молекулярного разнообразия этой структуры даже между штаммами эти результаты могут не отражать все разнообразие O-антигенов, представленное у немодельных видов. В настоящей работе проведен сравнительный анализ генов, участвующих в

синтезе O-антигена, для бактерий из семейства Oxalobacteraceae. В отличие от существующих исследований, преимущественно основанных на методе ПЦР, в нашей работе использован биоинформатический подход, а сравнение проведено не на уровне одиночных генов, а на уровне кластеров. Мы обнаружили, что в случае Oxalobacteraceae генетическая организация O-антигена представлена несколькими кластерами, находящимися на значительном удалении друг от друга в геноме бактерий. Наибольшее сходство кластеров наблюдалось внутри отдельных родов бактерий, что объясняется высокой изменчивостью O-антигенов. В работе описано сходство генов O-антигенов, присущее семейству в целом, а также рассмотрены отдельные уникальные случаи изменчивости их генетической структуры у отдельных бактерий.

Ключевые слова: кластеры генов O-антигена; гены липополисахарида; сравнительный анализ; O-антиген; Oxalobacteraceae; *Massilia*; *Collimonas*; *Janthinobacterium*; кластеры генов сахаридов.

## Introduction

The Oxalobacteraceae family belongs to the Burkholderiales order of Proteobacteria. According to the Integrated Taxonomic Information System ([www.itis.gov](http://www.itis.gov)) this family includes 55 verified species of 12 genera. Members of the Oxalobacteraceae family are stained negatively by Gram and presented in a wide range of habitats (outlined in Supplementary Materials, Table S1)<sup>1</sup>. Species were found in soils, including grassland, volcanic and heavy metal polluted soils, in water and glaciers (Baldani et al., 2014). Some of them are free-living, others may form various relationships with plants. Symbiotic species (*Massilia*, *Herbaspirillum*) are known to exhibit plant growth-promoting features, and can be beneficial in agriculture (Ofek et al., 2012; Peta et al., 2019; Grillo-Puertas et al., 2021). Occasionally, these relationships lead to plant diseases, for example, red stripe and mottle stripe diseases (Tuleski et al., 2020). The negative effect depends on the environment conditions. Examples of opportunistic features are described for *Janthinobacterium* and *Herbaspirillum* genera. Some species can be found in clinical samples and act as opportunistic pathogens for humans (Dhital et al., 2020).

Beneficial effects from Oxalobacteraceae bacteria are related to agriculture and medicine. Farming industry utilizes these bacteria to improve plant growth. Mutualistic bacteria facilitate nitrogen assimilation to increase crops productivity. In medicine, bacterial lipopolysaccharides (LPS) can be used for vaccine development. This modern medicine development is called glycoconjugate vaccines. The methodology is already verified on the members of Enterobacteraceae family (Bazhenova et al., 2021) and can be scaled to other bacteria. Beyond vaccines, information related to LPS lies in biosensor systems. Systems are able to identify bacteria in samples based on their LPS composition, in particular O-antigens (Sannigrahi et al., 2020).

O-antigen became a convenient feature for serotyping due to its variability. Diversity of the oligopolysaccharide units and the selection of the host immune system directed at them highly contribute to the variability of O-antigens. In addition to this selection, there is the bacteriophage effect on the bacterial cell (Xi et al., 2019). All these factors explain the emergence of different serotypes within the same bacterial species.

O-antigen is one part of bacterial LPS. Lipopolysaccharides are a specific structures (plural form) binding to the outer membrane of Gram-negative bacteria. It consists of three parts that are linked to each other in a particular order: phospholipid anchored to the membrane (lipid A or endotoxin), core region and O-antigen repeats. Lipid A is the hydrophobic domain an-

choring LPS in the membrane. In chemical structure, lipid A is a phospholipid based on glucosamine. It forms the monolayer of the outer membrane. Lipid A is responsible for the toxicity of Gram-negative bacteria. The second component of LPS is the core part. The first and the second LPS components are synthesized on the cytoplasmic side of the inner membrane of the bacterial cell, after which they are transported by ABC transporters into the periplasmic space (Valvano, 2015). The third component of LPS is O-antigen, which is synthesized separately from the previous parts. In a periplasmic space, all parts of LPS are combined together, then the fully synthesized LPS is transported to the outer leaflet of the cell membrane (Doerrler, 2006).

The composition of LPS and its parts varies between different species and between strains (Caroff, Karibian, 2003). In some strains O-antigen can be absent, thus referred to as “rough” LPS, others containing it are “smooth” (Erridge et al., 2002). The O-antigen consists of a series of repeating oligosaccharide units. The length and composition of the monomers vary quite widely among strains (Perepelov et al., 2009). Repeats can be homodimers or heterodimers. In addition, units can be linked linearly or can create a branched structure (Liu et al., 2020).

Sugar nucleotides are basic molecules that form an O-antigen backbone. The most common can be divided into several groups:

- dTDP-sugars (*rfb/rml* genes);
- CDP-sugars (*ddh* genes);
- GDP-sugars (*man* genes, *gmd*, *col*);
- UDP-glucoses (*ugd*, *gla*, *galE*);
- UDP-N-acetylglucosamines (*gne*, *gna*, *fnl* and *mna* genes).

Other nucleotide sugar genes include *nna* genes (N-acetylneuraminic acid synthesis), *hdd* genes and *gmh* (LD-mannoheptose and DD-mannoheptose) and *dmh* genes of 6-deoxy-D-mannoheptose synthesis pathway (Samuel, Reeves, 2003). The O-antigen chain is assembled via glycosyltransferases, which are responsible for combinations of sugar nucleotides.

The mechanisms of generating O-antigen and flipping are described in two variants: Wzy-dependent pathway and ABC-transporter pathway. The former is predominant among better-characterized O-antigens. A third variant is the synthase-dependent pathway. Unfortunately, it is poorly described and has been observed rarely, for instance, in *Salmonella* species (Kalynych et al., 2014).

The initiation of all O-antigen synthesis pathways is a transfer of a sugar monophosphate to the undecaprenyl phosphate (Und-P) molecule, resulting in sugar-pyrophosphate-undecaprenyl (sugar-Und-PP). Sugar-Und-PP is able to accept further glycosylation reactions (Kalynych et al., 2014).

<sup>1</sup> Supplementary Materials 1–6 are available in the online version of the paper: <https://doi.org/10.5281/zenodo.7410337>.

Uniquely to the Wzy-dependent pathway, Und-P-linked units are polymerized by Wzy (*wzy* gene) and subsequently flipped via Wzx (*wzx* gene). The chain length is controlled by Wzz protein (*wzz*). The completed structure is ligated to the outer core region via WaaL O-antigen ligase encoded by *waaL* (*rfaL*) gene (Han et al., 2012).

On the contrary, the ABC-transporter pathway needs only a single initiation reaction per O-antigen chain. Moreover, the entire polymerization process via glycosyltransferases is carried out in the cytoplasm. Then the completely generated O-antigen-Und-PP molecule is flipped to the periplasmic space by an ABC transporter, which is encoded by *wzt* and *wzm* genes. Similarly to the previously characterized pathway, the O-antigen ligase protein WaaL connects it to the core-lipid A (Samuel, Reeves, 2003).

In view of the above described, O-antigen becomes a highly variable structure. This feature makes the O-antigen attractive to a wide range of researchers. Nevertheless, there are rather few studies on comparative analysis of O-antigens and their genetic structure between bacteria at the family level. Most publications are devoted to single pathogenic or potentially pathogenic bacteria and avoid features of free-living or mutualistic species.

Detection and study of O-antigens have been made possible by the emergence of several methods involving both experimental and bioinformatics analysis of bacterial data. One of the traditional methods belonging to the first group is the bacterial glycotyping method based on the somatic antigen. In 2020, E.T. Sumrall et al. (2020) proposed a new method for quantitative separation of O-antigens. It is based on the use of a set of recombinant proteins that can interact with bacterial envelope receptors and domains. Bacterial O-antigens can also be detected by serological and agglutination test methods using sera specific to somatic antigens (Thakur et al., 2018). Another way to study O-antigen composition is the polymerase chain reaction method, which is widely used to compare O-antigens in several bacteria.

The emergence and subsequent decrease in the cost of sequencing opened new ways of O-antigen studying. *In silico* analysis methods have significantly reduced the time required for data processing, and many routine processes have been automated. Extensive databases have appeared that lead to the O-antigens analysis of several bacteria at once. In comparison to traditional methods of O-antigen detection, *in silico* methods are able to revise taxonomy misunderstandings, identify more genes related to O-antigen biosynthesis and evaluate their environment in a short time. Predicted features can be then verified by traditional laboratory methods. On the example of an Oxalobacteraceae member called *Janthinobacterium* sp. SLB01 (Belikov et al., 2021), the taxonomy was revised by this combined approach.

Here we present comparative analysis of O-antigens for 20 genomes from the Oxalobacteraceae family. According to the query in UniprotKB “(protein\_name: O-antigen) AND (taxonomy\_id:75682)” there are only 456 genes whose proteins are annotated as O-antigen biosynthesis genes for this family. Our bioinformatics approach based on homologues search eliminates difficulties in gene annotation. We also shift from describing single genes to comparing O-antigens at the level

of their candidate gene clusters to broad information about the gene content of Oxalobacteraceae O-antigens.

## Materials and methods

**Data.** Initial data was derived from NCBI databases and included 20 genomes. The main criterion of assembly selection was a rather high quality, that is, no more than ten contigs. The reason for such a criterion was to decrease the possibility of gene clusters being disrupted by unresolved sequences. Overall, we selected two *Collimonas* species (*C. arenae* and *C. fungivorans*), one species of genera *Herminiimonas* (*H. arsenitoxidans*), *Oxalobacter* (*O. formigenes*), and *Undibacterium* (*U. parvum*), four *Janthinobacterium* (*J. agaricidamnosum*, *J. lividum*, *J. svalbardensis*, *J. tructae*), two *Oxalicibacterium* (*O. faegigallinarum* and *O. flavum*) species and nine *Massilia* (*M. albidiflava*, *M. armeniaca*, *M. flava*, *M. oculi*, *M. plicata*, *M. putida*, *M. timonae*, *M. umbonata*, *M. violaceinigra*). Their RefSeq assembly accessions are presented in Supplementary Materials, Table S2.

**Quality control and annotation.** All 20 assemblies were additionally analyzed using QUAST tool, version 5.0.2 (Gurevich et al., 2013). The acceptable threshold number of contigs and scaffolds was eleven, only genomes with a lower number were selected. To obtain the most precise annotation, we used two annotation tools, Prokka version 1.14.6 (Seemann, 2014) and eggNOG version 2.1.6 (Huerta-Cepas et al., 2019).

**Putative O-antigen genes search.** Searching for genes coding components for O-antigen synthesis and processing based on their names was unproductive because of the abundance of various synonymous tags. Therefore, we used an approach based on orthology. All O-antigen related genes for *Escherichia coli* strains described in the paper (Iguchi et al., 2015) were obtained with their amino acid sequences and used as reference. We also added genes from the KEGG database, a pathway of O-antigen synthesis for *E. coli* <https://www.genome.jp/pathway/ecoi00541>. We additionally analyzed the O-antigen ligase gene *rfaL* (*waaL*), because it was shown that O-antigen may be absent in some bacteria (Kime et al., 2016). As *waaL* is essential for final stages of O-antigen processing for the majority of bacteria, its absence may be associated with a lack of OPS on the cell wall (Wang et al., 2010). This data consisted of gene sets for each serogroup and approximately 420 unique genes in total (Supplementary Materials, Table S3).

In order to find unique genes among this data, sequences were clustered using UCLUST (Edgar, 2010) algorithm with the usearch32 tool, with threshold identity > 0.4. The reason for the rather low threshold was the excessive amount of clusters at higher numbers, mainly because of high gene variation. For the next step, we chose centroids of each cluster as representative sequences.

To reveal genes that correspond to the processed O-antigen genes of *E. coli*, we used the tool Orthofinder (Emms, Kelly, 2019) (version 2.5.4), which is able to find orthogroups and orthologs. Centroids data was taken as a reference. We assign functions of *E. coli* reference genes to all the Oxalobacteraceae sequences that fall into the same orthologous group.

The gene cluster is defined as a set of genes involved in a common metabolic pathway located within the genomic region

of 27,000 bp in length (Cimermanic et al., 2014). However, another important parameter for our definition is genes on borders. Thus, for an array of three genes, if genes on the borders of the set coincide, we also define this set as a cluster. A more detailed investigation of the obtained gene clusters with respect to their structure, function and sequence similarity was conducted using eggNOG and BLAST (v.2.5.0+) tools.

Verification of the identified candidate genes was performed via functional Pfam domains search (Supplementary Materials, Table S4). Lists of domains were obtained manually, from (Iguchi et al., 2015; Pereira et al., 2015). The HMMER software hmmer.org version 3.3.2 allowed the detection of those domains in FASTA amino acid sequences of all genomes. Some genes were checked manually using online Pfam sequence search <https://pfam.xfam.org> (Mistry et al., 2021). Characterization of genes shown to be uninvolved in O-antigen biosynthesis processes was performed using KEGG databases (Kanehisa, 2000).

**Phylogenetic tree reconstruction.** Phylogenetic tree was constructed to explore evolutionary relationships between the chosen Oxalobacteraceae taxa. Several species of the Burkholderiaceae family were selected to create an outgroup (*Burkholderia sordidicola*, *B. unamae*, *B. symbiotica*, *Ralstonia pickettii*, *Cupriavidus necator*). 16S rRNA sequences for 13 Oxalobacteraceae species were derived from published papers (Lim et al., 2003; Caballero-Mellado et al., 2004; Zhang et al., 2006; Sheu et al., 2012; Baldani et al., 2014; Koh et al., 2017; Daniel et al., 2021; Jung et al., 2021). Barnmap version 0.9 (RRID:SCR\_015995) was used for seven other genomes (*C. arenae*, *C. fungivorans*, *J. agaricidamnosum*, *J. lividum*, *J. svalbardensis*, *M. timonae*, *O. flavum*) to derive 16S rRNA sequences (Supplementary Materials, Table S5).

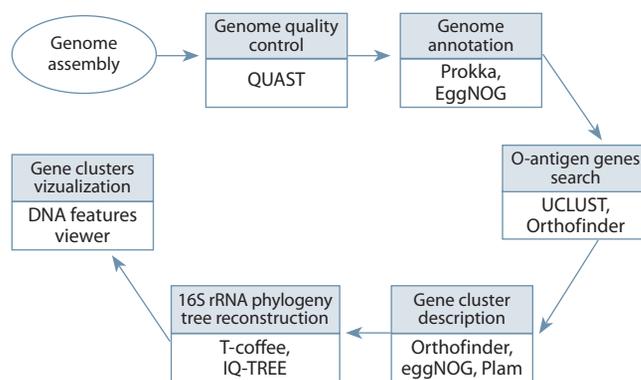
16S rRNA sequences were aligned using R-coffee, the T-coffee web-server RNA sequences alignment tool (Notredame et al., 2000). This tool takes into consideration the RNA secondary structure. Default multiple alignment options were chosen. The resulting alignment was used for constructing a phylogenetic tree using IQ-TREE web server (Nguyen et al., 2015). DNA was selected for sequence type, other options remained default. The best-fit model was TN+F+I+G4, the tree constructed with the Maximum likelihood method. Consensus tree was constructed from 1000 bootstrap trees and branch lengths were optimized by Maximum likelihood on original alignment. The results were visualized using Archaeopteryx 0.9928 (Han, Zmasek, 2009).

**Gene clusters visualization.** To visualize the found clusters we developed a Python script based on the DnaFeaturesViewer library (<https://edinburgh-genome-foundry.github.io/DnaFeaturesViewer/index.html#more-biology-software>). The code is available on this page [https://github.com/svetaafonnikova/O-antigen-project/blob/main/draw\\_cluster.py](https://github.com/svetaafonnikova/O-antigen-project/blob/main/draw_cluster.py). All steps of the data analysis algorithm are schematically depicted in Fig. 1.

## Results

### Assembly quality characterization

Out of all 20 assemblies, 15 were at the level of complete genomes. *M. timonae* assembly consisted of a single contig with N50 equal to the length of this contig. Two assemblies contained plasmid sequences (*M. putida* and *M. violacei-*



**Fig. 1.** Schematic representation of the data analysis algorithm used in the current study.

*nigra*). Another pair, *O. faecigallinarum* and *O. flavum*, contained ten and nine contigs, respectively.

Using IGV (v. 2.11.1) (Robinson et al., 2011) we confirmed that the identified O-antigen gene clusters were not located on plasmid fragments in case of plasmid containing genome assemblies. Secondly, O-antigen gene clusters were not situated on the borders of contigs, thus any breaks inside clusters were excluded.

### Description of gene clusters

In general, almost all of the analyzed species contained more than two O-antigen gene clusters. These clusters are scattered around the genome and include not only O-polysaccharide genes, but genes of other functions. The visualization for all 20 species can be found in Supplementary Materials, Fig. S1. In the text below, we will describe these clusters for each genus used in the study.

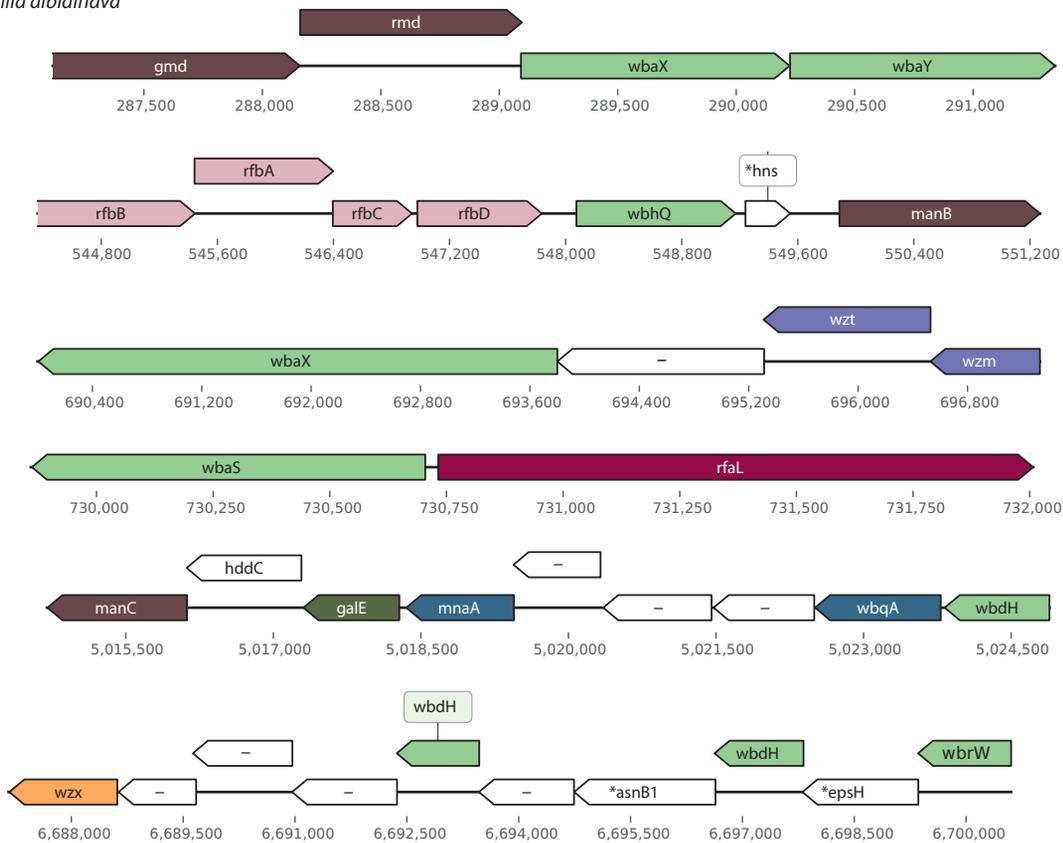
**Collimonas.** In both *C. arenae* and *C. fungivorans* we detected O-antigen ligase gene *rfaL* (or *waaL*) immediately adjacent to *galE* gene involved in nucleotide sugar synthesis. In addition, both genomes contain *wzm* and *wzt* genes. Furthermore, they share the same cluster with *manB* and *wfaK* on borders. All genes and their order coincide except one glycosyltransferase gene *wbaS*, absent in *C. fungivorans*.

Regarding other differences, the former species consists of three clusters, the latter consists of four. One of *C. fungivorans* clusters contains O-antigen unit synthesis (*rmd*, *gmd*, *manC*), processing genes (*wzm*, *wzt*) and a triplet of glycosyltransferase gene *wbaX*. Remarkably, in *C. arenae* these processing and unit synthesis genes are included in a single cluster with *rfaABCD* and *manB* genes on the borders.

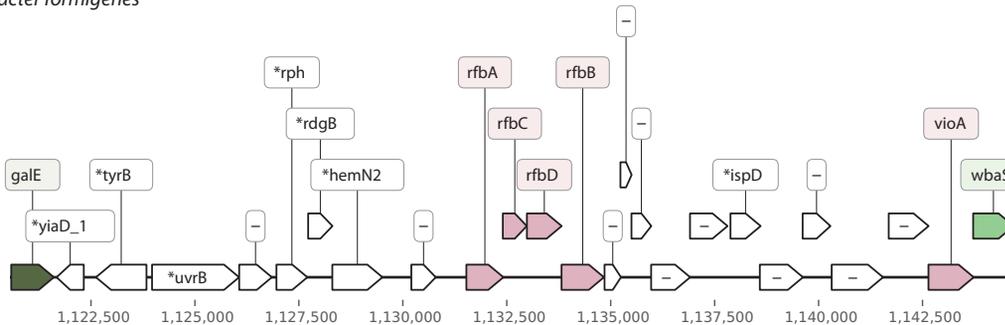
**Hermiimonas.** According to our analysis, *H. arsenitoxidans* genome possesses three O-antigen gene clusters, with *rfaL* gene located outside all of them without any OPS genes beside. Regarding genes involved in processing, only *wzx* was observed. There are duplication instances for L-Rhamnose biosynthesis gene *rfbD*, sugar transferase genes *wbaT* and *wbaS*. One cluster contains a rather small number of genes we are interested in compared to not O-antigen ones. These unnecessary for OPS production genes partake in phosphate metabolism.

**Janthinobacterium.** *J. lividum* carries two clusters and *J. agaricidamnosum* comprises three gene clusters involved

*Massilia albidiflava*



*Oxalobacter formigenes*



**Fig. 2.** The O-antigen gene clusters from *M. albidiflava* and *O. formigenes*. Unannotated genes are designated as “-”.

Colors code for biosynthetic pathways. Orange genes are involved in Wzx/Wzy pathway, violet are involved in Wzm/Wzt pathway, rose genes partake in dTDP-sugar pathway, dark green, in UDP-sugar pathway, brown, in GDP-sugar pathway, *rfaL* gene is coded in red. UDP-N-Acetylglucosamine genes are blue, and transferase genes are light green. Genes involved in other pathways are white. The complete graphical visualization of OPS gene clusters for other analyzed species can be found in Supplementary Materials, Fig. S1.

in O-antigen synthesis, whereas *J. svalbardensis* and *J. tructae* include four. The latter two share identical clusters with UDP-N-acetylglucosamine pathway, *wbqA* and *wbqB* on the one end and glycosyltransferase gene *wbdH* on the other. All but *J. agaricidamnosum* have duplications of *rfbABCD* genes. All four genes are duplicated in *J. lividum* and *J. svalbardensis*, *J. tructae* possesses three copies of *rfbA* and *rfbB*. Furthermore, the *J. tructae* cluster with *rfbBA* and *fnlA* borders is almost similar to a part of another larger O-antigen gene cluster. In *J. lividum* and *J. svalbardensis* we found a common OPS related gene cluster flanked by *wbqB* and *wbhQ*. This gene set includes dTDP-glucose pathway genes *rfbABCD* and *vioA*. Still, the latter species has glycosyltransferase *wbaS*

next to *wbqB*, which *J. lividum* lacks in this position. To add, *wzx* gene was located after *vioA* in *J. lividum*, however, we didn't observe any significant domains for *J. svalbardensis* in that position. The O-antigen ligase was observed in all genus members. It lies far from any depicted cluster.

Considering genes not included in our initial gene list, there are genes involved in LPS core synthesis (*waaD*), polysaccharide transport gene (*wza*), genes characteristic to O-antigen production in other bacteria species (*rfbG*, *rfbF*).

**Massilia.** According to our analysis, *Massilia* is the genus with the highest number of O-antigen gene clusters. *M. oculi* has six clusters, *M. flava*, *M. umbonata* and *M. violaceinigra* possess only four and others contain five clusters (Fig. 2).

We observed some patterns in gene clusters between species. All *Massilia* species carry the *gmd\_rmd\_wbaX\_wbaY* (in exact order) cluster. Only in *M. oculi* and *M. timonae* *rfbBDAC* genes (order in cluster) are presented as an independent cluster. In other genomes, these genes are surrounded by various O-antigen related genes. The same cluster with *rfbBACD* (order in cluster) genes and *manB* occurs in *M. violaceinigra*, *M. plicata*, *M. flava*, *M. armeniaca* and *M. albidiflava*. A single gene unrelated to O-antigen production is DNA-binding protein gene *hns*.

To add more similarity between *M. oculi* and *M. timonae*, they possess identical clusters consisting of *wbrW*, *wbdH*, *wbqB*, *ugd* on the one end and *wzx* on the other end. Genes located among them partake in infection initiation (*espH*), amino acid biosynthesis (*asnB1*), acyl-CoA and fatty acids biosynthesis (*fadD*).

In all assemblies, we observed *wzm* and *wzt* genes. Most of the species contain these genes in the order *wzm*, *wzt*, unannotated gene and *wbaX*. The group with such a set includes *M. albidiflava*, *M. armeniaca*, *M. oculi*, *M. putida*, *M. plicata*, and *M. timonae*. Another gene context is larger, the cluster is flanked by *wzm/wzt* and *vioA*. Between them are two glycosyltransferase genes *wbaX* with different lengths, unannotated genes and *gtrB*. The latter is a viral gene, and it can actually modify O-antigen structure. However, it was not described for the *E. coli* OPS gene cluster. Finally, in *M. violaceinigra* we found a unique set (not O-antigen biosynthesis gene cluster by our definition) of OPS processing genes and *wbaX*. There are three unannotated genes and two *wzt*. For the one beside *wzm* we didn't verify a specific domain, it was indicated as a gene not involved in O-antigen synthesis. The domain structure for *wzt* laying further from *wzt* was proved.

One of *M. armeniaca* clusters contains a full cluster described for *M. plicata*. It starts with *mnaA*, proceeds with *wbrW* and three copies of *wbdH*. In the former species, *wzx* with unannotated genes is added after the third *wbdH*. *M. umbonata* shares the most part of this cluster with *M. armeniaca*, except it lacks *mnaA* at the beginning. All genomes except *M. plicata* include the *galE*, *hddC* and *manC* part in the exact order in one cluster per genome.

Considering O-antigen ligase gene *rfaL*, in *M. albidiflava*, *M. oculi*, *M. plicata*, *M. timonae* and *M. violaceinigra* this gene is located next to *wbaS*. The rest of the species contain *rfaL* outside O-antigen clusters.

It can be noticed that some genes, for instance, *wbaS*, *rfaB* and *rfaB*, *mnaA*, *wbdH*, are presented in two or more copies in genomes.

**Oxalicibacterium.** Three clusters were identified for each species of the *Oxalicibacterium* genus. They share a cluster flanked by *wfaK* and *manC*. Their content slightly diverges from each other. *O. flavum* has more genes, including an additional O-antigen related gene *ugd*. The OPS ligase gene *rfaL* was identified in both assemblies, however, they are located in different contexts.

Oppositely to *O. flavum*, *O. faecigallinarum* carries UDP-N-Acetylglucosamine pathway genes (*fnlA*, *fnlB*, *mnaA*, *gne*, *wbqB*). On top of it, in the *O. faecigallinarum* we could locate duplications of the *rfaABCD* part, lying in discrete clusters and ordered in a different manner. However, *rfaD* gene in the

bigger cluster is rather dubious, the smaller length compared to other *rfaD* instances adds more uncertainty. We did not find this gene using Orthofinder analysis, although there is a Pfam domain corresponding to typical *rfaD* (Rm1D\_sub\_bind) and it was annotated as *rfaD* by EggNOG.

We could detect *wzt/wzm* genes only in *O. flavum* assembly. The second species probably either does not carry these genes or they can be located outside clusters in unread spaces between contigs.

**Oxalobacter.** For *O. formigenes* we identified a single OPS cluster carrying dTDP-sugar pathway genes *rfaABCD* and *vioA* and UDP-glucose synthesis gene *galE*. The rest of the genes in the cluster are involved in nucleotide metabolism and cofactor synthesis. Also, any O-antigen processing genes were undiscovered. We couldn't detect *rfaL* gene in the given assembly. Moreover, even NCBI databases don't have any information considering this gene or protein in the *Oxalobacter* genus.

**Undibacterium.** For *U. parvum* two clusters were identified, *wzt* and *wzm* genes, were located outside them. Interestingly, *wzt* gene is smaller in comparison to this gene's length in other Oxalobacteraceae species. Typical of them, *wzt* is longer than *wzm* by approximately 400 bp. In contrast, *U. parvum*'s *wzt* is almost the same size as *wzm*. Using Pfam service, the gene's domain (ABC\_tran) was verified.

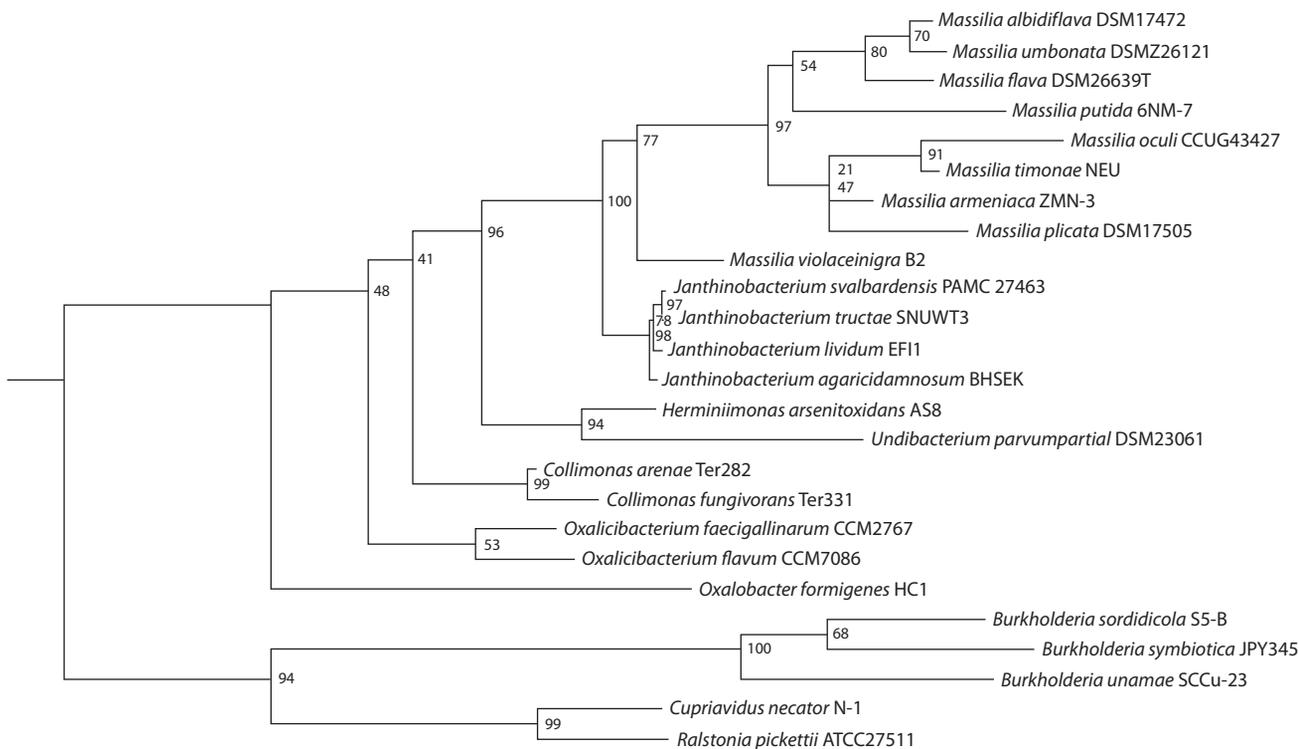
Both clusters possess transferase and nucleotide sugar genes. Most spaces between OPS synthesis genes are unannotated genes, except *dyp* (peroxidase) and *ansA* (asparaginase) genes. The cluster carrying *rfaABCD* genes has a copy of *manC* gene and two *wbaX* genes, which have different sizes.

### Phylogenetic tree

The phylogenetic dendrogram based on 16S rRNA showed that the chosen species clustered together considering their genera (Fig. 3). There had been no study including all species and their exact strains used in the current work. Therefore, we could compare only some clades of the tree. Similar to other studies, the first species to branch off is *Oxalobacter* species. Contrary to literature reports, our tree has a distinct *Oxalicibacterium* group and *Collimonas* with the rest of the species of the Oxalobacteraceae family (Baldani et al., 2014). However, the bootstrap support is rather small at this node. The *Janthinobacterium* group formation coincided with other papers (Jung et al., 2021). Some *Massilia* species clustered according to literature (Feng et al., 2016; Ren et al., 2018). Also, we obtained an unresolved node between *M. armeniaca* and *M. plicata*. The gene lengths used in the analysis varied between 1400 and 1500 bp for most cases (see Supplementary Materials, Table S5).

### Discussion

In this work, we determined candidate genes involved in O-antigen biosynthesis in bacteria from the Oxalobacteraceae family. In comparison to well-studied *E. coli* O-antigen genes, they are presented in the form of several clusters. A similar situation has already been described for non-model bacteria (Hug et al., 2010). These clusters are dispersed across the genome. Clusters include O-antigen genes together with additional genes, which are necessary for LPS biosynthesis (for example, for core part synthesis and LPS parts binding) or



**Fig. 3.** Phylogenetic reconstruction of Oxalobacteraceae family members selected for the study based on 16S rRNA and created using Maximum likelihood method.

The consensus tree was obtained from 1000 bootstrap trees. The sequence data is described in Supplementary Materials, Table S3.

partake in other processes. The *E. coli* O-antigen gene cluster was studied by traditional laboratory methods, in particular, by PCR (DebRoy et al., 2011; Iguchi et al., 2015). These methods aim to detect specific genes, whereas *in silico* methods take into account the gene environment. In other words, they allow structures to be studied at the cluster level. Thus, our approach helps to expand understanding of the O-antigen genetic composition in bacterial genomes.

During OPS genetic structure comparison, we identified common features presented in all species inside the Oxalobacteraceae family. In particular, the group of *rfbABCD* genes was detected in each bacterium. The order of these genes varies, however, they are always placed together in one cluster. No one gene has deletions, nonsense mutations and other sequence abnormalities. According to the results, the studied bacteria should have a correct dTDP-rhamnose synthesis.

More similarities were found within each genus. These similarities relate mainly to individual genes or pairs of genes. A possible explanation lies in the high level of variability of O-antigens and the rate of bacterial mutations. O-antigens undergo changes so frequently that most of the similarities occur at the species or lower levels rather than at the genus or family level (Liu et al., 2008).

In 13 bacteria species, *wzm* and *wzt* genes were detected. We consider the Wzm-Wzt transporters pathway as the main path of O-antigen biosynthesis in this case (Wang et al., 2010). Wzx-Wzy pathway was not confirmed due to the absence of *wzy* genes.

Another interesting finding concerns gene duplication. The most repetitive genes were identified in *Massilia* species (see

Fig. 2). Its O-antigens clusters may contain up to three copies of the same gene. We suggest two possible explanations. The first one is related to the biological features of LPS. The same gene can provide the synthesis of several parts of LPS. The appearance of additional gene copies can increase the amount of protein in the cell or maintain its level in case one of the gene copies is broken. The second explanation is linked with an algorithm of O-antigen genes search. In our approach, genes are detected according to the principle of homology, so similar genes can be assigned the same name.

Symbiotic bacteria *Oxalobacter formigenes* lacks O-antigen ligase gene (*walL*) in O-antigen clusters, which may indicate the absence of O-antigen. The lack of the mentioned structure was discussed by J.K. Kim et al. (2016) for *Burkholderia* bacteria species. With our results, we confirm the possibility of loss of O-antigen genes in symbiotic bacterial species.

## Conclusion

Overall, the findings of this study indicate differences of non-model bacteria from the model one by the example of the Oxalobacteraceae family. We suggest that the characterized OPS gene cluster composition is atypical. So far, most papers, which explored these genes for other bacteria, described only a single gene cluster. The O-antigen genetics of non-model bacteria is highly diverse, which is proved by the bioinformatic approach. The search for homologous sequences allows us to expand and deepen our understanding of gene clusters involved in O-antigen biosynthesis. Further investigation of the Oxalobacteraceae O-antigen genetic composition can be confirmed by laboratory methods.

## References

- Baldani J.I., Rouws L., Cruz L.M., Olivares F.L., Schmid M., Hartmann A. The family *Oxalobacteraceae*. In: Rosenberg E., DeLong E.F., Lory S., Stackebrandt E., Thompson F. (Eds.). *The Prokaryotes*. Berlin; Heidelberg: Springer, 2014;919-974. DOI 10.1007/978-3-642-30197-1\_291.
- Bazhenova A., Gao F., Bolgiano B., Harding S.E. Glycoconjugate vaccines against *Salmonella enterica* serovars and *Shigella* species: existing and emerging methods for their analysis. *Biophys. Rev.* 2021;13(2):221-246. DOI 10.1007/s12551-021-00791-z.
- Belikov S.I., Petrushin I.S., Chernogor L.I. Genome analysis of the *Janthinobacterium* sp. strain SLB01 from the diseased sponge of the *Lubomirskia baicalensis*. *Curr. Issues Mol. Biol.* 2021;43(3):2220-2237. DOI 10.3390/cimb43030156.
- Caballero-Mellado J., Martínez-Aguilar L., Paredes-Valdez G., Estrada-de los Santos P. *Burkholderia unamae* sp. nov., an N<sub>2</sub>-fixing rhizospheric and endophytic species. *Int. J. Syst. Evol. Microbiol.* 2004; 54(4):1165-1172. DOI 10.1099/ijs.0.02951-0.
- Caroff M., Karibian D. Structure of bacterial lipopolysaccharides. *Carbohydr. Res.* 2003;338(23):2431-2447. DOI 10.1016/j.carres.2003.07.010.
- Cimermancic P., Medema M.H., Claesen J., Kurita K., Wieland Brown L.C., Mavrommatis K., Pati A., Godfrey P.A., Koehrsen M., Clardy J., Birren B.W., Takano E., Sali A., Linington R.G., Fischbach M.A. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell.* 2014;158(2):412-421. DOI 10.1016/j.cell.2014.06.034.
- Daniel S.L., Moradi L., Paiste H., Wood K.D., Assimos D.G., Holmes R.P., Nazzal L., Hatch M., Knight J. Forty years of *Oxalobacter formigenes*, a gutsy oxalate-degrading specialist. *Appl. Environ. Microbiol.* 2021;87(18):e0054421. DOI 10.1128/AEM.00544-21.
- DebRoy C., Roberts E., Fratamico P.M. Detection of O antigens in *Escherichia coli*. *Anim. Heal. Res. Rev.* 2011;12(2):169-185. DOI 10.1017/S1466252311000193.
- Dhital R., Paudel A., Bohra N., Shin A.K. *Herbaspirillum* infection in humans: a case report and review of literature. *Case Rep. Infect. Dis.* 2020;2020:9545243. DOI 10.1155/2020/9545243.
- Doerler W.T. Lipid trafficking to the outer membrane of Gram-negative bacteria. *Mol. Microbiol.* 2006;60(3):542-552. DOI 10.1111/j.1365-2958.2006.05130.x.
- Edgar R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 2010;26(19):2460-2461. DOI 10.1093/bioinformatics/btq461.
- Emms D.M., Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20(1):238. DOI 10.1186/s13059-019-1832-y.
- Erridge C., Bennett-Guerrero E., Poxton I.R. Structure and function of lipopolysaccharides. *Microbes Infect.* 2002;4(8):837-851. DOI 10.1016/S1286-4579(02)01604-0.
- Feng G.-D., Yang S.Z., Li H.P., Zhu H.H. *Massilia putida* sp. nov., a dimethyl disulfide-producing bacterium isolated from wolfram mine tailing. *Int. J. Syst. Evol. Microbiol.* 2016;66(1):50-55. DOI 10.1099/IJSEM.0.000670.
- Grillo-Puertas M., Villegas J.M., Pankiewicz V.C.S., Tadra-Sfeir M.Z., Teles Mota F.J., Hebert E.M., Brusamarello-Santos L., Pedraza R.O., Pedrosa F.O., Rapisarda V.A., Souza E.M. Transcriptional responses of *Herbaspirillum seropedicae* to environmental phosphate concentration. *Front. Microbiol.* 2021;12:666277. DOI 10.3389/FMICB.2021.666277.
- Gurevich A., Saveliev V., Vyahhi N., Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29(8):1072-1075. DOI 10.1093/bioinformatics/btt086.
- Han M.V., Zmasek C.M. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics.* 2009;10:356. DOI 10.1186/1471-2105-10-356.
- Han W., Wu B., Li L., Zhao G., Woodward R., Pettit N., Cai L., Thon V., Wang P.G. Defining function of lipopolysaccharide O-antigen ligase WaaL using chemoenzymatically synthesized substrates. *J. Biol. Chem.* 2012;287(8):5357-5365. DOI 10.1074/jbc.M111.308486.
- Huerta-Cepas J., Szklarczyk D., Heller D., Hernández-Plaza A., Forslund S.K., Cook H., Mende D.R., Letunic I., Rattei T., Jensen L.J., von Mering C., Bork P. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 2019;47(D1):D309-D314. DOI 10.1093/nar/gky1085.
- Hug I., Couturier M.R., Rooker M.M., Taylor D.E., Stein M., Feldman M.F. *Helicobacter pylori* lipopolysaccharide is synthesized via a novel pathway with an evolutionary connection to protein N-glycosylation. *PLoS Pathog.* 2010;6(3):e1000819. DOI 10.1371/journal.ppat.1000819.
- Iguchi A., Iyoda S., Kikuchi T., Ogura Y., Katsura K., Ohnishi M., Hayashi T., Thomson N.R. A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA Res.* 2015;22(1):101-107. DOI 10.1093/dnares/dsu043.
- Integrated Taxonomic Information System (ITIS). On-line database. www.itis.gov. (Retrieved 08.21.2022). CC0. https://doi.org/10.5066/F7KH0KKB.
- Jung W.J., Kim S.W., Giri S.S., Kim H.J., Kim S.G., Kang J.W., Kwon J., Lee S.B., Oh W.T., Jun J.W., Park S.C. *Janthinobacterium tractae* sp. nov., isolated from kidney of rainbow trout (*Oncorhynchus mykiss*). *Pathogens.* 2021;10(2):229. DOI 10.3390/pathogens10020229.
- Kalynych S., Morona R., Cygler M. Progress in understanding the assembly process of bacterial O-antigen. *FEMS Microbiol. Rev.* 2014; 38(5):1048-1065. DOI 10.1111/1574-6976.12070.
- Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27-30. DOI 10.1093/nar/28.1.27.
- Kim J.K., Park H.Y., Lee B.L. The symbiotic role of O-antigen of *Burkholderia* symbiont in association with host *Riptortus pedestris*. *Dev. Comp. Immunol.* 2016;60:202-208. DOI 10.1016/j.dci.2016.02.009.
- Koh H.-W., Hur M., Kang M.-S., Ku Y.-B., Ghai R., Park S.-J. Physiological and genomic insights into the lifestyle of arsenite-oxidizing *Herminiimonas arsenitoxidans*. *Sci. Rep.* 2017;7(1):15007. DOI 10.1038/s41598-017-15164-4.
- Lim Y.W., Baik K.S., Han S.K., Kim S.B., Bae K.S. *Burkholderia sordidicola* sp. nov., isolated from the white-rot fungus *Phanerochaete sordida*. *Int. J. Syst. Evol. Microbiol.* 2003;53(5):1631-1636. DOI 10.1099/ijs.0.02456-0.
- Liu B., Furevi A., Perepelov A.V., Guo X., Cao H., Wang Q., Reeves P.R., Knirel Y.A., Wang L., Widmalm G. Structure and genetics of *Escherichia coli* O antigens. *FEMS Microbiol. Rev.* 2020;44(6):655-683. DOI 10.1093/femsre/fuz028.
- Liu B., Knirel Y.A., Feng L., Perepelov A.V., Senchenkova S.N., Wang Q., Reeves P.R., Wang L. Structure and genetics of *Shigella* O antigens. *FEMS Microbiol. Rev.* 2008;32(4):627-653. DOI 10.1111/J.1574-6976.2008.00114.X.
- Mistry J., Chuguransky S., Williams L., Qureshi M., Salazar G.A., Sonnhammer E.L.L., Tosatto S.C.E., Paladin L., Raj S., Richardson L.J., Finn R.D., Bateman A. Pfam: The protein families database in 2021. *Nucleic Acids Res.* 2021;49(D1):D412-D419. DOI 10.1093/nar/gkaa913.
- Nguyen L.T., Schmidt H.A., Von Haeseler A., Minh B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 2015;32(1):268-274. DOI 10.1093/MOLBEV/MSU300.
- Notredame C., Higgins D.G., Heringa J. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 2000; 302(1):205-217. DOI 10.1006/JMBI.2000.4042.
- Ofek M., Hadar Y., Minz D. Ecology of root colonizing *Massilia* (Oxalobacteraceae). *PLoS One.* 2012;7(7):e40117. DOI 10.1371/journal.pone.0040117.
- Pereira S.B., Mota R., Vieira C.P., Vieira J., Tamagnini P. Phylum-wide analysis of genes/proteins related to the last steps of assembly and

- export of extracellular polymeric substances (EPS) in cyanobacteria. *Sci. Rep.* 2015;5(1):14835. DOI 10.1038/srep14835.
- Perepelov A.V., Liu B., Senchenkova S.N., Shashkov A.S., Feng L., Wang L., Knirel Y.A. Structure of O-antigen and functional characterization of O-antigen gene cluster of *Salmonella enterica* O47 containing ribitol phosphate and 2-acetimidoylamino-2,6-dideoxy-L-galactose. *Biochemistry (Moscow)*. 2009;74(4):416-420. DOI 10.1134/S0006297909040099.
- Peta V., Raths R., Bücking H. Draft genome sequence of *Massilia* sp. strain ONC3, a novel bacterial species of the Oxalobacteraceae family isolated from garden soil. *Microbiol. Resour. Announc.* 2019; 8(32):e00377-19. DOI 10.1128/MRA.00377-19.
- Ren M., Li X., Zhang Y., Jin Y., Li S., Huang H. *Massilia armeniaca* sp. nov., isolated from desert soil. *Int. J. Syst. Evol. Microbiol.* 2018; 68(7):2319-2324. DOI 10.1099/IJSEM.0.002836.
- Robinson J.T., Thorvaldsdóttir H., Winckler W., Guttman M., Lander E.S., Getz G., Mesirov J.P. Integrative genomics viewer. *Nat. Biotechnol.* 2011;29(1):24-26. DOI 10.1038/nbt.1754.
- Samuel G., Reeves P. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr. Res.* 2003;338(23):2503-2519. DOI 10.1016/j.carres.2003.07.009.
- Sannigrahi S., Arumugasamy S.K., Mathiyarasu J., K.S. Magnetosome-anti-*Salmonella* antibody complex based biosensor for the detection of *Salmonella typhimurium*. *Mater. Sci. Eng. C. Mater. Biol. Appl.* 2020;114:111071. DOI 10.1016/j.msec.2020.111071.
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30(14):2068-2069. DOI 10.1093/bioinformatics/btu153.
- Sheu S.Y., Chou J.H., Bontemps C., Elliott G.N., Gross E., James E.K., Sprent J.I., Young J.P.W., Chen W.M. *Burkholderia symbiotica* sp. nov., isolated from root nodules of *Mimosa* spp. native to north-east Brazil. *Int. J. Syst. Evol. Microbiol.* 2012;62(9):2272-2278. DOI 10.1099/IJS.0.037408-0.
- Sumrall E.T., Röhrig C., Hupfeld M., Selvakumar L., Du J., Dunne M., Schmelcher M., Shen Y., Loessner M.J. Glycotyping and specific separation of *Listeria monocytogenes* with a novel bacteriophage protein tool kit. *Appl. Environ. Microbiol.* 2020;86(13):e00612-20. DOI 10.1128/AEM.00612-20.
- Thakur N., Jain S., Changotra H., Shrivastava R., Kumar Y., Grover N., Vashist J. Molecular characterization of diarrheagenic *Escherichia coli* pathotypes: Association of virulent genes, serogroups, and antibiotic resistance among moderate-to-severe diarrhea patients. *J. Clin. Lab. Anal.* 2018;32(5):e22388. DOI 10.1002/jcla.22388.
- Tuleski T.R., Kimball J., do Amaral F.P., Pereira T.P., Tadra-Sfeir M.Z., de Oliveira Pedrosa F., Maltempi de Souza E., Balint-Kurti P., Monteiro R.A., Stacey G. *Herbaspirillum rubrisubalbicans* as a phyto-pathogenic model to study the immune system of *Sorghum bicolor*. *Mol. Plant Microbe Interact.* 2020;33(2):235-246. DOI 10.1094/MPMI-06-19-0154-R.
- Valvano M.A. Genetics and biosynthesis of lipopolysaccharide. In: Tang Y.-W., Sussman M., Liu D., Poxton I., Schwartzman J. (Eds.). *Molecular Medical Microbiology*. Academic Press, 2015;55-89. DOI 10.1016/B978-0-12-397169-2.00004-4.
- Wang L., Wang Q., Reeves P.R. The variation of o antigens in gram-negative bacteria. *Subcell. Biochem.* 2010;53:123-152. DOI 10.1007/978-90-481-9078-2\_6.
- Xi D., Wang X., Ning K., Liu Q., Jing F., Guo X., Cao B. O-antigen gene clusters of *Plesiomonas shigelloides* serogroups and its application in development of a molecular serotyping scheme. *Front. Microbiol.* 2019;10:741. DOI 10.3389/FMICB.2019.00741.
- Zhang Y.Q., Li W.J., Zhang K.Y., Tian X.P., Jiang Y., Xu L.H., Jiang C.L., Lai R. *Massilia dura* sp. nov., *Massilia albidiflava* sp. nov., *Massilia plicata* sp. nov. and *Massilia lutea* sp. nov., isolated from soils in China. *Int. J. Syst. Evol. Microbiol.* 2006;56(2):459-463. DOI 10.1099/IJS.0.64083-0.

#### ORCID ID

S.D. Afonnikova [orcid.org/0000-0001-7969-8015](https://orcid.org/0000-0001-7969-8015)  
A.S. Komissarov [orcid.org/0000-0001-6981-7316](https://orcid.org/0000-0001-6981-7316)  
P.D. Kuchur [orcid.org/0000-0002-9415-577X](https://orcid.org/0000-0002-9415-577X)

**Acknowledgements.** The authors thank the Collective Center of ICG SB RAS "Bioinformatics" Joint Computational Center for the use of computational resources.

**Conflict of interest.** The authors proclaim that the research was conducted in the absence of any financial relationship that could be interpreted as a potential conflict of interest.

Received August 29, 2022. Revised October 10, 2022. Accepted October 26, 2022.

Перевод на английский язык <https://vavilov.elpub.ru/jour>

## Контекстные сигналы в митохондриальных микроРНК млекопитающих

О.В. Вишневыский<sup>1</sup>, П.С. Ворожейкин<sup>2</sup>✉, И.И. Титов<sup>1, 2, 3</sup>

<sup>1</sup> Федеральныи исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>2</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

<sup>3</sup> Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия

✉ pavel.vorozheykin@gmail.com

**Аннотация.** МикроРНК – это малые некодирующие РНК, которые регулируют экспрессию генов на пост-транскрипционном уровне в цитоплазме, и, таким образом, играют важную роль в большом числе биологических процессов. Последние исследования обнаружили присутствие последовательностей микроРНК не только в цитоплазме, но и внутри митохондрий. Такие микроРНК (так называемые митомиры, mitomiRs) могут иметь ядерное или митохондриальное происхождение, при этом для некоторых из них установлена роль в регуляции функций митохондриальных генов, а для большинства – она пока неизвестна. Выявление нуклеотидных сигналов, уникальных для митомиров, может помочь определить эту роль. В нашей работе составлена выборка экспериментально обнаруженных митомиров человека, мыши и крысы. С целью выделения сигналов, которые могут быть ответственны за функционирование митомиров и за их транспортировку в митохондрии или из них, осуществлен контекстный анализ для полученных последовательностей митомиров. Для трех видов в группе данных митомиры/не-митомиры и в группе всех микроРНК из базы miRBase выявлены статистически перепредставленные 8-буквенные мотивы (уровень значимости  $p < 0.01$ , с учетом поправки Бонферрони на множественность сравнения). Для этих мотивов обнаружены закономерности их локализации в функционально значимых участках для разных типов микроРНК. Для рассматриваемой группы митомиры/не-митомиры также обнаружены статистически значимые особенности нуклеотидного состава последовательностей микроРНК возле границ разрезания комплексами Drosha/Dicer (критерий независимости  $\chi^2$  Пирсона для первых трех позиций микроРНК с уровнем значимости  $p < 0.05$ ). Наблюдаемые частоты нуклеотидов, предположительно, могут указывать на наличие у митомиров (в сравнении с не-митомирами) более однородного разрезания прай-миРНК комплексом Drosha при формировании 5'-конца последовательностей. Результаты работы могут быть полезными для выявления сигналов, принимающих участие в возникновении, процессинге и функциях митомиров. Ключевые слова: микроРНК; пре-миРНК; митомир; митохондрия.

**Для цитирования:** Вишневыский О.В., Ворожейкин П.С., Титов И.И. Контекстные сигналы в митохондриальных микроРНК млекопитающих. *Вавиловский журнал генетики и селекции*. 2022;26(8):819-825. DOI 10.18699/VJGB-22-99

## The context signals of mitochondrial miRNAs (mitomiRs) of mammals

O.V. Vishnevsky<sup>1</sup>, P.S. Vorozheykin<sup>2</sup>✉, I.I. Titov<sup>1, 2, 3</sup>

<sup>1</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup> Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia

✉ pavel.vorozheykin@gmail.com

**Abstract.** MicroRNAs (miRNAs) are small non-coding RNAs that regulate gene expression at the post-transcriptional level in the cytoplasm and play an important role in a wide range of biological processes. Recent studies have found that the miRNA sequences are presented not only in the cytoplasm, but also in the mitochondria. These miRNAs (the so-called mitomiRs) may be the sequences of nuclear or mitochondrial origin; some of them are involved in regulation of the mitochondrial gene functions, while the role of others is still unknown. The identification of nucleotide signals, which are unique to mitomiRs, may help to determine this role. We formed a dataset that combined the experimentally discovered mitomiRs in human, rat and mouse. To isolate signals that may be responsible for the mitomiRs' functions or for their translocation from or into mitochondria a context analysis was carried out for the sequences. For three species in the group mitomiRs/non-mitomiRs and the group of all miRNAs from the miRBase database statistically overrepresented 8-letter motifs were identified ( $p$ -value  $< 0.01$  with Bonferroni correction for multiple comparisons), for these motifs the patterns of the localization in functionally important regions for different types of miRNAs were found. Also, for the group mitomiRs/non-mitomiRs we found the statistically significant features of the miRNA nucleotide context near the Dicer and Drosha cleavage

sites (Pearson's  $\chi^2$  test of independence for the first three positions of the miRNA,  $p$ -value < 0.05). The observed nucleotide frequencies may indicate a more homogeneous pri-miRNA cleavage by the Drosha complex during the formation of the 5' end of mitomiRs. The obtained results can help to determine the role of the nucleotide signals in the origin, processing, and functions of the mitomiRs.

Key words: miRNA; pre-miRNA; mitomiR; mitochondrion.

**For citation:** Vishnevsky O.V., Vorozheykin P.S., Titov I.I. The context signals of mitochondrial miRNAs (mitomiRs) of mammals. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):819-825. DOI 10.18699/VJGB-22-99

## Введение

К настоящему времени достаточно хорошо исследованы основные пути биогенеза микроРНК, которые начинаются в ядре клетки и заканчиваются в цитоплазме (Bartel, 2018). При изучении нуклеотидного состава микроРНК и их предшественников (прайм-/пре-миРНК) установлено наличие сигналов, которые могут влиять на функции микроРНК и на процесс созревания микроРНК на отдельных стадиях биогенеза. Нуклеотидная последовательность микроРНК может как непосредственно определять функции микроРНК, так и оказывать влияние на точность разрезания комплексами процессинга Drosha/Dicer 5'-конца микроРНК и, таким образом, формировать сайт-специфически модифицированные микроРНК со сдвигом так называемого seed region (Starega-Roslan et al., 2015a, b; Rolle et al., 2016) (seed region – это участок со 2-го по 7-й нуклеотид микроРНК, который отвечает за ее адресацию).

Наличие мотивов в одноцепочечных концах (UG; CNNC) или в основании (CUC/GHG) прайм-миРНК, в терминальной петле (GU) шпильки пре-миРНК может приводить к блокированию или, наоборот, к содействию процессинга микроРНК (Auyeung et al., 2013; Fang, Bartel, 2015; Nguyen et al., 2015; Starega-Roslan et al., 2015a, b; Rolle et al., 2016; Vorozheykin, Titov, 2020). Помимо ядра и цитоплазмы, эти малые РНК-последовательности, а также белки комплексов для их процессинга обнаружены в органеллах, например в митохондриях (Kren et al., 2009; Bandiera et al., 2011; Wang et al., 2015). Эти наблюдения свидетельствуют как о вероятном существовании новых путей биогенеза микроРНК внутри митохондрий, так и о возможной транспортировке зрелых микроРНК между цитоплазмой и митохондриями пока неизвестными транспортными комплексами. Существование таких митохондриальных микроРНК (так называемых митомиров, mitomiRs) поднимает вопросы об их эволюционном происхождении и их функциях внутри и вне органелл. Присутствуют ли у них структурные особенности, обеспечивающие их функции и транспортировку внутрь или из митохондрий?

В настоящей статье собрана информация из опубликованных работ об экспериментально наблюдаемых микроРНК в митохондриях (митомиров). Для отобранных митомиров выполнена оценка контекстных особенностей последовательностей на предмет возможного влияния нуклеотидных сигналов на происхождение, процессинг и функции митомиров.

## Материалы и методы

В работе использованы последовательности микроРНК *Homo sapiens*, *Mus musculus*, *Rattus norvegicus* из базы данных miRBase (<http://miRBase.org>, выпуск 22.1) (Kozo-

mara et al., 2019). Общее количество микроРНК для трех рассматриваемых видов составляет 5398 последовательностей.

Информация о митомирах получена из статей с экспериментальным изучением (методами RT-qPCR, microarray, qRT-PCR) локализации микроРНК внутри и вне митохондрий для разных организмов и тканей (Kren et al., 2009; Bian et al., 2010; Bandiera et al., 2011; Barrey et al., 2011; Mercer et al., 2011; Das et al., 2012; Sripada et al., 2012; Wang et al., 2015). По этим работам для человека, мыши и крысы сформированы две выборки последовательностей: митомиры (652 последовательности микроРНК, наблюдаемые в митохондриях) и все остальные микроРНК из базы miRBase (мы будем называть их не-митомиры, 4766 последовательностей).

Для изучения особенностей последовательностей разных групп микроРНК выполнен поиск статистически перепредставленных (уровень значимости  $p < 0.01$ , с учетом поправки Бонферрони на множественность сравнения) олигонуклеотидных мотивов с помощью программы ARGO (Vishnevsky, Kolchanov, 2005) для *de novo* поиска мотивов в 15-буквенном коде для группы выборок митомиры/не-митомиры и группы всех микроРНК трех рассматриваемых видов из базы miRBase. При поиске мотивов для всех микроРНК из базы miRBase программа ARGO оценивает ожидаемую долю случайных последовательностей с мононуклеотидным частотным составом как у анализируемой выборки, содержащих мотив по случайным причинам.

Для полученных мотивов выполнена оценка совпадения мотивов в рамках каждой из рассмотренных групп и между двумя группами. Для каждой пары мотивов определяется коэффициент сходства (мера) Jaccard, в нашем случае  $\frac{N_{\text{совпад.}}}{N_{\text{общее}}}$ , где  $N_{\text{совпад.}}$  – число 4-буквенных нуклеотидных последовательностей, которые совпадают у двух мотивов;  $N_{\text{общее}}$  – число всех 4-буквенных нуклеотидных последовательностей, соответствующих обоим мотивам. Данная мера принимает значение от 0 до 1, где значение 0 демонстрирует полное различие, а значение 1 – полное совпадение двух мотивов.

Для оценки вероятности получения коэффициента Jaccard по случайным причинам использован метод, предложенный в работе (Real, Vargas, 1996), в которой случайное значение коэффициента Jaccard предполагается распределенным по биномиальному закону (с точностью до нормировки). Для выявленных мотивов определена локализация мотивов в последовательности микроРНК и анализ нуклеотидного состава с целью обнаружения неоднородности вырезания микроРНК из предшественника комплексами Drosha и Dicer. При анализе локализации

мотивов для группы всех микроРНК из базы miRBase в качестве «контрастной» выборки использованы случайные позиции, выбранные внутри последовательностей микроРНК изучаемой выборки.

### Результаты и обсуждение

При обработке публикаций с экспериментально наблюдаемыми микроРНК в митохондриях нами выявлено упоминание 652 уникальных идентификаторов микроРНК. При этом 272 последовательности из найденных микроРНК (митомиров) можно охарактеризовать как высокодостоверные, так как они или верифицированы дополнительно методами RT-qPCR/qRT-PCR, или в данных microarray-экспериментов наблюдаются в большем количестве внутри митохондрий, нежели вне их.

Отдельно нужно отметить семь митомиров, последовательности которых полностью присутствуют в митохондриальном геноме человека: hsa-miR-1974, hsa-miR-1977, hsa-miR-1978, hsa-miR-4461, hsa-miR-4463, hsa-miR-4484, hsa-miR-4485-3p, что может служить дополнительным подтверждением их достоверности. При этом ввиду наложения последовательности микроРНК на tРНК митохондрий к текущему моменту из базы miRBase удалены упоминания следующих микроРНК: hsa-miR-1974,

hsa-miR-1977, hsa-miR-1978. Из базы также удалена микроРНК hsa-miR-4461; экспериментально полученные для нее данные не удовлетворяют требованиям, предъявляемым при аннотации микроРНК. Таким образом, из базы miRBase были исключены последовательности, которые не соответствуют текущим известным путям биогенеза микроРНК, но могут формироваться неизвестными неканоническими путями.

Для последующего изучения и сравнения характеристик митомиров в работе также использована выборка не-митомиров общим количеством 4766 последовательностей. В нее вошли все микроРНК человека, мыши и крысы из рассматриваемой базы miRBase, за исключением митомиров.

С помощью программы ARGO (Vishnevsky, Kolchanov, 2005) проанализированы группы выборок микроРНК: группа из двух выборок митомиров/не-митомиров и группа всех микроРНК из базы miRBase. Для каждой из групп отобрано 40 (табл. 1) и 44 (табл. 2) 8-нуклеотидных мотивов, которые имеют статистически значимое различие встречаемости в выборках микроРНК в каждой из групп (уровень значимости  $p < 0.01$ , с учетом поправки Бонферрони на множественность сравнения). Для мотивов внутри каждой из групп, а также для мотивов из разных

**Таблица 1.** Мотивы, имеющие статистически значимое различие встречаемости в выборках последовательностей митомиров и не-митомиров

№ п/п	Мотив	% митомиров	% не-митомиров	Уровень значимости, $p$	№ п/п	Мотив	% митомиров	% не-митомиров	Уровень значимости, $p$
1	CAKTSNAN	8.44	0.87	$3.8 \times 10^{-26}$	21	MWCMBAVH	9.82	2.86	$9.7 \times 10^{-8}$
2	KTGCANDK	8.90	1.26	$5.6 \times 10^{-21}$	22	RKTGYWBH	11.81	3.98	$1.2 \times 10^{-7}$
3	HASHWSBD	28.53	11.89	$5.8 \times 10^{-21}$	23	GYHSHBDG	18.10	7.97	$1.7 \times 10^{-7}$
4	WMAGKGCD	6.29	0.54	$1.8 \times 10^{-20}$	24	YWCMCTBT	5.21	0.85	$3.8 \times 10^{-7}$
5	MNTVCANK	13.96	3.40	$3.6 \times 10^{-20}$	25	KKVAACMH	5.98	1.17	$8.9 \times 10^{-7}$
6	HRVRNTSH	34.97	17.14	$1.4 \times 10^{-18}$	26	CTNVRBTS	9.66	3.13	$2.1 \times 10^{-5}$
7	KBAGGTWG	5.21	0.41	$8.5 \times 10^{-17}$	27	CTRKNBVW	14.88	6.34	$2.4 \times 10^{-5}$
8	AGSAVCWY	5.21	0.41	$8.7 \times 10^{-17}$	28	RCABCMHN	6.13	1.40	$5.7 \times 10^{-5}$
9	RHASHWSB	20.86	7.93	$9.1 \times 10^{-16}$	29	YCMYWMMM	6.29	1.48	$7.5 \times 10^{-5}$
10	RCADTSDH	9.97	2.13	$7.9 \times 10^{-15}$	30	SAGVAMHN	8.13	2.45	$2.0 \times 10^{-4}$
11	RSTRDIT	8.13	1.44	$4.5 \times 10^{-14}$	31	WKMYCMKA	5.21	1.06	$2.1 \times 10^{-4}$
12	WMDWSCWB	15.49	5.08	$1.2 \times 10^{-13}$	32	NMYASDGS	10.43	3.79	$3.5 \times 10^{-4}$
13	HSVYDGDN	44.02	26.58	$1.9 \times 10^{-12}$	33	KGARNMCI	5.52	1.22	$4.0 \times 10^{-4}$
14	WRMACWTB	6.13	0.85	$3.5 \times 10^{-12}$	34	TSRGWSDG	5.98	1.46	$9.8 \times 10^{-4}$
15	CCHKBWGD	9.36	2.20	$2.3 \times 10^{-11}$	35	WCCHBTHS	6.60	1.77	$1.3 \times 10^{-3}$
16	GBYWVYWG	12.12	3.63	$6.5 \times 10^{-11}$	36	SAVWSSCW	6.13	1.59	$2.9 \times 10^{-3}$
17	KGYNASW	10.74	3.03	$4.8 \times 10^{-10}$	37	STRHDGTT	5.06	1.11	$3.5 \times 10^{-3}$
18	CADKGNTD	8.13	1.79	$5.8 \times 10^{-10}$	38	NGGCWMDS	7.06	2.07	$4.1 \times 10^{-3}$
19	GWGSTNVY	9.66	2.60	$4.2 \times 10^{-9}$	39	HCBRRCT	5.37	1.26	$5.3 \times 10^{-3}$
20	WCAKSWR	6.44	1.24	$3.5 \times 10^{-8}$	40	YSTSRSTS	5.98	1.55	$5.8 \times 10^{-3}$

Примечание. % митомиров – доля последовательностей митомиров, содержащих мотив; % не-митомиров – доля последовательностей не-митомиров, содержащих мотив.

Здесь и в табл. 2: в качестве результата отобраны мотивы с уровнем значимости  $p < 0.01$  (с учетом поправки Бонферрони на множественность сравнения).

**Таблица 2.** Мотивы, имеющие статистически значимое различие встречаемости в выборке всех микроРНК базы miRBase в сравнении со случайными последовательностями с тем же мононуклеотидным составом, что и анализируемая выборка микроРНК

№ п/п	Мотив	% miRBase	% случайных	Уровень значимости, <i>p</i>	№ п/п	Мотив	% miRBase	% случайных	Уровень значимости, <i>p</i>
1	YNCKBYCB	12.09	5.36	$1.1 \times 10^{-71}$	23	HRHABYRC	5.63	2.99	$2.8 \times 10^{-15}$
2	BYNCYKYC	11.26	4.94	$2.0 \times 10^{-67}$	24	GCKSVKBK	6.28	3.46	$3.1 \times 10^{-15}$
3	AWRYRHWY	6.33	2.10	$1.3 \times 10^{-59}$	25	KTGYABDD	5.63	3.01	$1.0 \times 10^{-14}$
4	RHARHRHW	11.79	5.90	$1.0 \times 10^{-50}$	26	GTWDWHYV	5.15	2.76	$9.9 \times 10^{-13}$
5	NCKKYCBV	11.09	5.45	$1.2 \times 10^{-49}$	27	RHBTKTGH	5.94	3.36	$1.8 \times 10^{-12}$
6	WDYAYDKW	9.21	4.16	$2.1 \times 10^{-49}$	28	HWYVYAYR	6.11	3.49	$3.0 \times 10^{-12}$
7	RHAWWYRY	5.06	1.63	$4.2 \times 10^{-48}$	29	NRRMRSSA	8.40	5.29	$4.4 \times 10^{-12}$
8	VGGMDVNG	11.92	6.10	$4.8 \times 10^{-48}$	30	RVKGGMRV	7.88	4.88	$5.0 \times 10^{-12}$
9	YRTANANV	5.04	1.86	$3.9 \times 10^{-37}$	31	CBKCYCNV	5.76	3.33	$2.2 \times 10^{-10}$
10	NHYVVCAG	9.47	4.81	$8.4 \times 10^{-37}$	32	KCCNBKBC	5.89	3.48	$2.0 \times 10^{-9}$
11	BTBYCYKY	9.21	4.66	$5.7 \times 10^{-36}$	33	HATHNYWY	5.70	3.35	$3.1 \times 10^{-9}$
12	YKHCTYYH	7.97	3.89	$2.5 \times 10^{-33}$	34	NKGWTDTH	5.06	2.89	$9.3 \times 10^{-9}$
13	ANBGHWDH	16.11	10.11	$4.4 \times 10^{-33}$	35	ASDHAVWW	5.37	3.15	$2.6 \times 10^{-8}$
14	CDGKVNNN	38.28	30.07	$7.4 \times 10^{-29}$	36	BCDGTKHY	5.30	3.11	$5.5 \times 10^{-8}$
15	RRMDGNAR	9.63	5.33	$5.0 \times 10^{-28}$	37	WGDRMHKG	8.99	6.10	$1.0 \times 10^{-7}$
16	GRGRHDGD	9.10	4.94	$6.7 \times 10^{-28}$	38	WWWTYRBD	5.41	3.27	$7.6 \times 10^{-7}$
17	DYAYDGTN	6.02	2.82	$6.2 \times 10^{-26}$	39	TBTMMYHY	5.30	3.30	$5.1 \times 10^{-5}$
18	WHAYAHNS	6.24	3.07	$2.0 \times 10^{-23}$	40	KSRGNBAG	6.31	4.12	$6.0 \times 10^{-5}$
19	NSDTNTHT	9.10	5.32	$1.5 \times 10^{-20}$	41	HMCMKYCH	5.44	3.50	$6.9 \times 10^{-4}$
20	TVYNYVCA	6.39	3.38	$1.3 \times 10^{-18}$	42	GWSGVDMN	7.88	5.54	$1.9 \times 10^{-3}$
21	DRYBTKTG	5.35	2.72	$1.0 \times 10^{-16}$	43	GWGHKBAB	5.08	3.26	$4.1 \times 10^{-3}$
22	TGBRRWKW	5.70	2.98	$1.3 \times 10^{-16}$	44	TWVTDWRH	5.19	3.37	$9.7 \times 10^{-3}$

Примечание. % miRBase – доля последовательностей базы miRBase, содержащих мотив; % случайных – оцененная программой ARGO доля случайных последовательностей, содержащих данный мотив по случайным причинам.

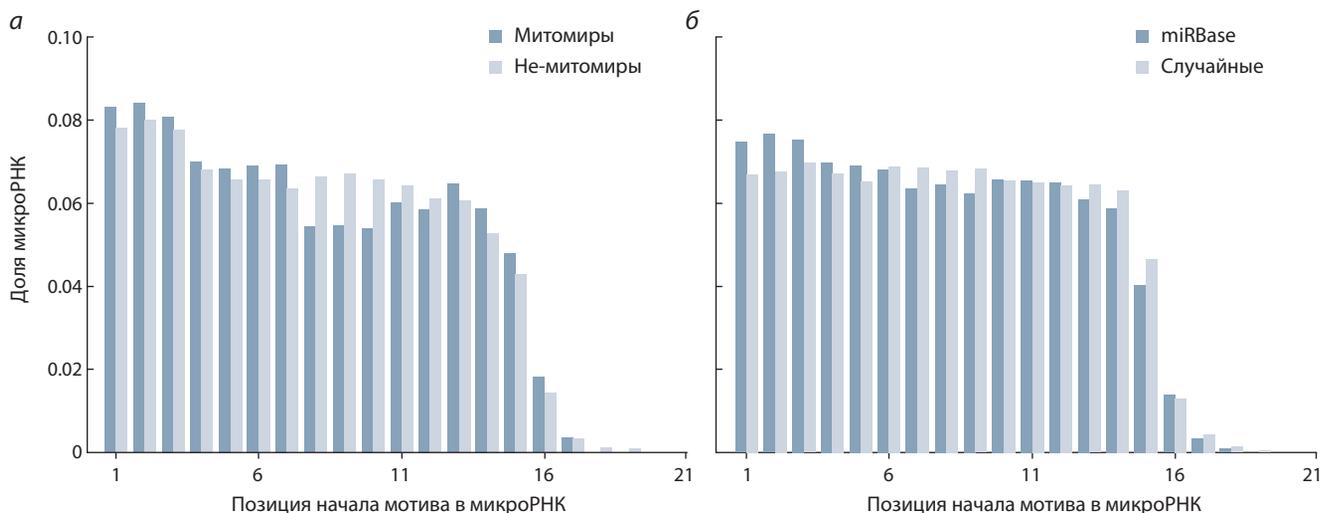
групп среднее значение коэффициента сходства Jaccard (исключая из расчета нулевые значения) для всех трех вычислений не превышает 0.02.

Для двух мотивов, KTGCANDK из группы митомиров/не-митомиров и KTGYABDD из группы всех микроРНК, отмечаются максимальный коэффициент (0.3) и минимальная вероятность наблюдения такого коэффициента по случайным причинам (0.81). Данные мотивы встречаются в 193 последовательностях из группы митомиров/не-митомиров и в 315 последовательностях группы всех микроРНК из базы miRBase. За исключением этих двух мотивов, коэффициент Jaccard для всех остальных рассматриваемых пар мотивов не превышает 0.13, вероятность наблюдения соответствующих коэффициентов по случайным причинам менее 0.001. Эти данные свидетельствуют о низкой степени совпадения мотивов как внутри каждой из групп, так и между двумя группами.

Такие различия нуклеотидного состава между выборками митомиров и не-митомиров могут выступать в качестве специфических сигналов для процессинга митомиров, например для распознавания и перемещения

последовательностей в/из митохондрий транспортными комплексами или для реализации специфических функций митомиров через прямое связывание с мишенями в митохондриальной или клеточной ДНК. В то же время мотивы, обнаруженные в выборке всех микроРНК, могут соответствовать сигналам, общим для процессинга и функций микроРНК, независимо от их локализации.

Для обеих рассматриваемых групп выборок первая позиция мотивов имеет тенденцию располагаться в начале микроРНК (рис. 1): максимальная доля последовательностей наблюдается для мотивов с началом в позициях 1–3 микроРНК. В полученных наблюдениях присутствует статистически значимая зависимость типа последовательностей микроРНК от позиций локализации начала мотива (критерий независимости  $\chi^2$  Пирсона, уровни значимости *p*:  $4.46 \times 10^{-2}$  и  $6.58 \times 10^{-5}$  соответственно для группы митомиров/не-митомиров и для группы все микроРНК из базы miRBase/случайные позиции в микроРНК). При этом для митомиров, в отличие от остальных выборок, наблюдается значительное сокращение числа микроРНК, в которых начало мотива располагается в по-



**Рис. 1.** Доля микроРНК в зависимости от позиций начала мотивов, встречающихся в последовательностях митомиров и не-митомиров (а) и во всех микроРНК из базы miRBase (б).

Для каждой микроРНК с мотивом определяется начальная с 5'-конца позиция мотива в последовательности. Если один мотив встречается в микроРНК несколько раз или несколько мотивов встречаются в одной микроРНК, каждый случай рассматривается независимо и порождает структуру данных (микроРНК, позиция мотива). График нормирован на общее количество полученных структур по всем мотивам. Спад наблюдений в позициях с номерами больше 15 обусловлен вариабельностью длин последовательностей микроРНК от 15 до 28 нуклеотидов.

зиях 8–10 микроРНК. Возможная причина в том, что для всех микроРНК (и митомиров, и не-митомиров) так называемый seed region является наиболее консервативным и значимым участком с функциональной точки зрения, поэтому рассматриваемые 8-буквенные консервативные мотивы часто «охватывают» эту область.

В противовес этому мотивы, начинающиеся с позиций 8–10 в микроРНК, располагаются в области так называемого дополнительного seed region (~13–16), который, предположительно, менее консервативен в митомирах и реже участвует в процессе связывания микроРНК с мишенью по сравнению с не-митомирами (см. рис. 1, а). Выборка всех микроРНК большей частью состоит из не-митомиров, поэтому результаты наблюдений для нее примерно совпадают с результатами наблюдений для не-митомиров.

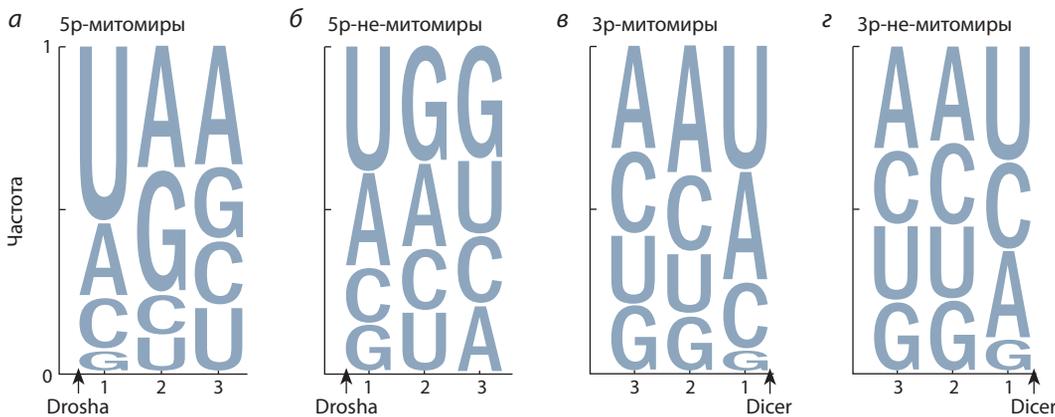
Для обнаруженных мотивов отмечаются разные паттерны локализации внутри последовательностей микроРНК. Один мотив может быть как в нескольких разных микроРНК с разным местоположением внутри последовательностей (например, мотив KTGCA<sub>NDK</sub> с уровнем значимости  $p = 5.6 \times 10^{-21}$  встречается с позиции 14 с 5'-конца митомира hsa-miR-92a-1-5p и с позиции 2 с 5'-конца митомира mmi-miR-19b-3p), так и внутри одной микроРНК, в том числе не пересекаясь друг с другом (например, для митомира hsa-miR-33a-5p мотив KTGCA<sub>NDK</sub> встречается дважды, начиная с позиции 1 и с позиции 12 с 5'-конца).

Таким образом, вариабельность локализации мотивов в микроРНК может свидетельствовать как о функциональной важности этих нуклеотидных сигналов для отдельных микроРНК, так и о возможном участии сигналов в процессинге микроРНК, в частности в отборе и перемещении последовательностей митомиров между митохондрией и цитоплазмой.

Для рассматриваемых групп микроРНК отмечается увеличение доли последовательностей, в которых мотивы начинаются в позициях 1–3 (см. рис. 1), поэтому далее мы осуществили контекстный анализ этого участка только для тех последовательностей митомиров и не-митомиров, в которых ранее были обнаружены мотивы. Для первых трех позиций 5'-конца 5p- и 3p-микроРНК вычислены попозиционные частоты встречаемости нуклеотидов (рис. 2).

Для митомиров из 5p-ветви пре-миРНК в первой позиции преимущественно наблюдается U и совсем редко – G, при этом в позиции 2 преимущественно встречается A или G (см. рис. 2, а). У не-митомиров отмечается увеличение в первой позиции числа G и A и уменьшение числа U (см. рис. 2, б). Для сайта разрезания Drosha наблюдается инверсия в позициях 2–3 между G в не-митомирах и A – в митомирах. В каждой из трех позиций частоты нуклеотидов демонстрируют зависимость от типа микроРНК (критерий независимости  $\chi^2$  Пирсона, уровни значимости  $p$ :  $2.89 \times 10^{-31}$ ,  $1.03 \times 10^{-28}$ ,  $1.79 \times 10^{-42}$  для позиций 1, 2 и 3 соответственно), при этом позиция 3 демонстрирует наиболее значимое различие частот между типами микроРНК, в отличие от позиций 1 и 2.

Сравнивая наблюдаемый нуклеотидный состав 5'-концов микроРНК для митомиров/не-митомиров с результатами исследования точности разрезания пре-миРНК комплексами Drosha и Dicer из работы (Starega-Roslan et al., 2015b), можно предположить, что для митомиров из 5p-ветви пре-миРНК разрезание Drosha осуществляется более точно, нежели для не-митомиров. То есть для митомиров формируются более однородный 5'-конец и соответствующий ему так называемый seed region, что может свидетельствовать о большей консервативности функций митомиров в сравнении с не-митомирами. Обнаруженные сигналы однородности разрезания для



**Рис. 2.** Частоты встречаемости нуклеотидов первых трех позиций начиная с 5'-конца микроРНК из 5p- и 3p-ветвей пре-миРНК для выборок митомиров (а, в) и не-митомиров (б, г), в последовательностях которых обнаружены мотивы.

Размеры букв пропорциональны частотам встречаемости нуклеотидов. По оси X отображаются номера позиций в микроРНК, отсчет от 5'-конца. Стрелками показаны области разрезания комплексами Dicer или Drosha. Для позиций 1–3 в 5p-микроРНК и для позиций 1–2 в 3p-микроРНК присутствует статистически значимая зависимость типа последовательностей микроРНК от частот встречаемости нуклеотидов в рассматриваемой позиции (критерий независимости  $\chi^2$  Пирсона, уровни значимости  $p$ :  $2.89 \times 10^{-31}$ ,  $1.03 \times 10^{-28}$ ,  $1.79 \times 10^{-42}$ ,  $1.17 \times 10^{-9}$ ,  $3.23 \times 10^{-10}$  для соответствующих позиций).

митомиров могут говорить как о возможном существовании в митохондриях более точного Drosha-подобного комплекса для процессинга микроРНК, так и о вероятном компенсировании неточности разрезания нуклеотидным составом последовательностей прай-миРНК, выбираемых для процессинга комплексом Drosha.

Для 5p-не-митомиров контекст смещается в сторону неоднородного разрезания (т. е. более активной сайт-специфической модификации микроРНК), и в таком случае не-митомиров могут выступать в качестве фактора функциональной изменчивости. Можно предположить, что митохондрии не «терпят» вариабельности «своих» микроРНК и в процессе эволюции устранили изоформы регуляторных последовательностей, а наблюдаемые митомиров могут являться оставшимися консервативными последовательностями, ведущими свое происхождение со времен доместикации предков митохондрий.

Для митомиров и не-митомиров из 3p-ветви пре-миРНК не наблюдается заметных отличий в нуклеотидном составе, за исключением инверсии в позиции 1 второго и третьего по популярности нуклеотидов (см. рис. 2, в и г). Статистически значимую зависимость позиционных нуклеотидных частот от типа микроРНК демонстрируют только позиции 1 и 2 (критерий независимости  $\chi^2$  Пирсона, уровни значимости  $p$ :  $1.17 \times 10^{-9}$  и  $3.23 \times 10^{-10}$  соответственно). Сравнение наблюдаемых частот нуклеотидов с результатами статьи (Starega-Roslan et al., 2015b) не дает однозначных выводов о качестве разрезания 5'-границы 3p-микроРНК комплексом Dicer.

### Заключение

В настоящей работе сформирована выборка экспериментально подтвержденных митомиров и произведен нуклеотидный анализ их последовательностей. Для группы данных митомиров/не-митомиров и для группы всех микроРНК из базы miRBase найдены статистически перепредставленные 8-буквенные мотивы внутри последо-

вательностей микроРНК. Эти мотивы демонстрируют, что последовательности митомиров могут представлять новый, неканонический класс микроРНК. Мотивы для группы митомиров/не-митомиров могут выступать в качестве сигналов для процессинга митомиров (например, участие при перемещении митомиров в/из митохондрий), для реализации функций митомиров (через связывание с мишенями в мтДНК или клеточной ДНК), в то время как мотивы группы всех микроРНК могут соответствовать сигналам, общим для процессинга и функций микроРНК, независимо от их локализации в клетке.

Нуклеотидный состав митомиров (в сравнении с не-митомирами) возле 5'-границ разрезания комплексами Drosha/Dicer, предположительно, может свидетельствовать о более однородном формировании 5'-конца последовательностей митомиров и, следовательно, о более консервативном функционале этих последовательностей.

### Список литературы / References

- Auyeung V.C., Ulitsky I., McGeary S.E., Bartel D.P. Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell*. 2013;152(4):844-858. DOI 10.1016/j.cell.2013.01.031.
- Bandiera S., Rüberg S., Girard M., Cagnard N., Hanein S., Chrétien D., Munnich A., Lyonnet S., Henrion-Caude A. Nuclear outsourcing of RNA interference components to human mitochondria. *PLoS One*. 2011;6(6):e20746. DOI 10.1371/journal.pone.0020746.
- Barrey E., Saint-Auret G., Bonnamy B., Damas D., Boyer O., Gidrol X. Pre-microRNA and mature microRNA in human mitochondria. *PLoS One*. 2011;6(5):e20220. DOI 10.1371/journal.pone.0020220.
- Bartel D.P. Metazoan microRNAs. *Cell*. 2018;173(1):20-51. DOI 10.1016/j.cell.2018.03.006.
- Bian Z., Li L.-M., Tang R., Hou D.-X., Chen X., Zhang C.-Y., Zen K. Identification of mouse liver mitochondria-associated miRNAs and their potential biological functions. *Cell Res*. 2010;20(9):1076-1078. DOI 10.1038/cr.2010.119.
- Das S., Ferlito M., Kent O.A., Fox-Talbot K., Wang R., Liu D., Raghavachari N., Yang Y., Wheelan S.J., Murphy E., Steenbergen C.

- Nuclear miRNA regulates the mitochondrial genome in the heart. *Circ. Res.* 2012;110(12):1596-1603. DOI 10.1161/CIRCRESAHA.112.267732.
- Fang W., Bartel D.P. The menu of features that define primary microRNAs and enable *de novo* design of microRNA genes. *Mol. Cell.* 2015;60(1):131-145. DOI 10.1016/j.molcel.2015.08.015.
- Kozomara A., Birgaoanu M., Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 2019;47(D1):D155-D162. DOI 10.1093/nar/gky1141.
- Kren B.T., Wong P.Y.-P., Sarver A., Zhang X., Zeng Y., Steer C.J. MicroRNAs identified in highly purified liver-derived mitochondria may play a role in apoptosis. *RNA Biol.* 2009;6(1):65-72. DOI 10.4161/rna.6.1.7534.
- Mercer T.R., Neph S., Dinger M.E., Crawford J., Smith M.A., Shearwood A.-M.J., Haugen E., Bracken C.P., Rackham O., Stamatoyannopoulos J.A., Filipovska A., Mattick J.S. The human mitochondrial transcriptome. *Cell.* 2011;146(4):645-658. DOI 10.1016/j.cell.2011.06.051.
- Nguyen T.A., Jo M.H., Choi Y.-G., Park J., Kwon S.C., Hohng S., Kim V.N., Woo J.-S. Functional anatomy of the human microprocessor. *Cell.* 2015;161(6):1374-1387. DOI 10.1016/j.cell.2015.05.010.
- Real R., Vargas J. M. The probabilistic basis of Jaccard's index of similarity. *Syst. Biol.* 1996;45(3):380-385. DOI 10.1093/sysbio/45.3.380.
- Rolle K., Piwecka M., Belter A., Wawrzyniak D., Jeleniewicz J., Barciszewska M.Z., Barciszewski J. The sequence and structure determine the function of mature human miRNAs. *PLoS One.* 2016; 11(3):e0151246. DOI 10.1371/journal.pone.0151246.
- Sripada L., Tomar D., Prajapati P., Singh R., Singh A.K., Singh R. Systematic analysis of small RNAs associated with human mitochondria by deep sequencing: detailed analysis of mitochondrial associated miRNA. *PLoS One.* 2012;7(9):e44873. DOI 10.1371/journal.pone.0044873.
- Starega-Roslan J., Galka-Marciniak P., Krzyzosiak W.J. Nucleotide sequence of miRNA precursor contributes to cleavage site selection by Dicer. *Nucleic Acids Res.* 2015a;43(22):10939-10951. DOI 10.1093/nar/gkv968.
- Starega-Roslan J., Witkos T., Galka-Marciniak P., Krzyzosiak W. Sequence features of Drosha and Dicer cleavage sites affect the complexity of isomiRs. *Int. J. Mol. Sci.* 2015b;16(12):8110-8127. DOI 10.3390/ijms16048110.
- Vishnevsky O.V., Kolchanov N.A. ARGO: a web system for the detection of degenerate motifs and large-scale recognition of eukaryotic promoters. *Nucleic Acids Res.* 2005;33(Web Server Iss.):W417-W422. DOI 10.1093/nar/gki459.
- Vorozheykin P.S., Titov I.I. Erratum to: How animal miRNAs structure influences their biogenesis. *Russ. J. Genet.* 2020;56(8):1012-1024. DOI 10.1134/S1022795420220019.
- Wang W.-X., Visavadiya N.P., Pandya J.D., Nelson P.T., Sullivan P.G., Springer J.E. Mitochondria-associated microRNAs in rat hippocampus following traumatic brain injury. *Exp. Neurol.* 2015;265:84-93. DOI 10.1016/j.expneurol.2014.12.018.

---

#### ORCID ID

I.I. Titov [orcid.org/0000-0002-2691-3292](https://orcid.org/0000-0002-2691-3292)

**Благодарности.** Работа ОВВ и ИИТ выполнена при финансовой поддержке бюджетного проекта № FWNR-2022-0020. Работа ИТ поддержана Курчатовским геномным центром ИЦиГ СО РАН (075-15-2019-1662).

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 07.09.2022. После доработки 17.11.2022. Принята к публикации 17.11.2022.

Перевод на английский язык <https://vavilov.elpub.ru/jour>

## Свойства малого мира научных организаций определяют динамику публикационной активности в области миРНК

А.Б. Фирсов<sup>1</sup> , И.И. Титов<sup>2, 3</sup>

<sup>1</sup> Институт систем информатики им. А.П. Ершова Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>2</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>3</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

 [artyomfirsov@mail.ru](mailto:artyomfirsov@mail.ru)

**Аннотация.** Многие научные статьи стали доступны в цифровом виде, что позволяет запрашивать данные статей и, в частности, автоматически собирать метаданные, включая данные об аффилиации. Это, в свою очередь, можно использовать для количественных оценок научной области, например для идентификации организаций и анализа графа соавторства этих организаций для извлечения базовой структуры науки. В настоящей работе рассмотрена область исследования микроРНК, а именно граф соавторства организаций и анализ его эволюции. Чтобы решить проблему вариативности написания названия организаций, был предложен алгоритм сортировки логических векторов признаков *k-mer/n-gram*. В нем используется тот факт, что содержание аффилиации довольно консистентно для одной и той же организации. Для учета ошибок написания и других артефактов названия организации в поле метаданных аффилиации наш подход преобразует упоминание организации внутри аффилиации в *K-Mer (n-gram)* булевый вектор присутствия. Далее векторы всех аффилиаций из набора данных лексикографически сортируются, образуя группы упоминаемых организаций. Таким подходом был кластеризован набор данных аффилиаций в области исследования микроРНК и определены названия уникальных организаций, что позволило построить граф соавторства на уровне научных организаций. С помощью этого графа показано, что рост области исследования микроРНК контролируется архитектурой малого мира сети научных организаций и испытывает степенной рост с показателем степени  $2.64 \pm 0.23$  для числа организаций в соответствии с диаметром сети, предлагая модель роста новых научных направлений. Скорость публикации первой статьи по микроРНК у организации при ее взаимодействии с другой организацией, уже публиковавшейся в этой области, аппроксимируется как  $0.184 \pm 0.002 \text{ год}^{-1}$ .

Ключевые слова: *k-mer*; *n-gram*; миРНК; электронная библиотека; соавторство организаций; малый мир.

**Для цитирования:** Фирсов А.Б., Титов И.И. Свойства малого мира научных организаций определяют динамику публикационной активности в области миРНК. *Вавиловский журнал генетики и селекции*. 2022;26(8):826-829. DOI 10.18699/VJGB-22-100

## Small world of the miRNA science drives its publication dynamics

A.B. Firsov<sup>1</sup> , I.I. Titov<sup>2, 3</sup>

<sup>1</sup> A.P. Ershov Institute of Informatics Systems of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>2</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

 [artyomfirsov@mail.ru](mailto:artyomfirsov@mail.ru)

**Abstract.** Many scientific articles became available in the digital form which allows for querying articles data, and specifically the automated metadata gathering, which includes the affiliation data. This in turn can be used in the quantitative characterization of the scientific field, such as organizations identification, and analysis of the co-authorship graph of those organizations to extract the underlying structure of science. In our work, we focus on the miRNA science field, building the organization co-authorship network to provide the higher-level analysis of scientific community evolution rather than analyzing author-level characteristics. To tackle the problem of the institution name writing variability, we proposed the *k-mer/n-gram* boolean feature vector sorting algorithm, KOFER in short. This approach utilizes the fact that the contents of the affiliation are rather consistent for the same organization, and to account for writing errors and other organization name variations within the affiliation metadata field, it converts the organization mention within the affiliation to the *K-Mer (n-gram)* Boolean presence vector. Those vectors for all affiliations in the dataset are further lexicographically sorted, forming groups of organization mentions. With that approach, we clustered the miRNA field affiliation dataset and extracted unique

organization names, which allowed us to build the co-authorship graph on the organization level. Using this graph, we show that the growth of the miRNA field is governed by the small-world architecture of the scientific institution network and experiences power-law growth with exponent  $2.64 \pm 0.23$  for organization number, in accordance with network diameter, proposing the growth model for emerging scientific fields. The first miRNA publication rate of an organization interacting with already publishing organization is estimated as  $0.184 \pm 0.002 \text{ year}^{-1}$ .

Key words: k-mer; n-gram; miRNA; digital library; organization co-authorship; small world.

**For citation:** Firsov A.B., Titov I.I. Small world of the miRNA science drives its publication dynamics. *Vavilovskii Zhurnal Genetiki i Selektii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):826-829. DOI 10.18699/VJGB-22-100

## Введение

Научные структуры стимулируют продуктивность научной работы, обеспечивая исследователей материально-техническими условиями и научной средой. Один из факторов эффективности научной работы – взаимодействие ученых в виде обмена идеями или совместной работы, которое проявляется в виде соавторства научных публикаций. Анализ соавторства исследовательских институтов, а не характеристик на уровне авторов дает возможность обеспечить более высокий уровень анализа эволюции научного сообщества, в частности организацию из «невидимых колледжей» или развитие международного сотрудничества в глобальном масштабе (Leydesdorff et al., 2013). Подобные исследования направлены на поиск причин конкуренции и сотрудничества в конкретных областях (Wagner, Leydesdorff, 2005), а также на выявление закономерностей международной публикационной активности (Ribeiro et al., 2018). В целом для понимания структуры научного сообщества и процесса распространения знаний в области науки анализ должен проводиться как на уровне авторов, так и на уровне организаций.

Граф имеет свойство малого мира, если  $L \propto \log(N)$ , где  $L$  – среднее кратчайшее расстояние графа,  $N$  – количество вершин графа. Другими словами, любые две вершины достижимы из другой посредством малого количества переходов через другие вершины, но при этом вероятность того, что они смежные, мала.

Такой тип сетей встречается во многих явлениях: например, распространение инфекции (Liu et al., 2015), нейронные связи (Muldoon et al., 2016) и др. Отдельный интерес представляет анализ эффекта малого мира в распространении знаний (Shi, Guan, 2016), и поэтому в нашей работе проверяется, является ли граф взаимодействия организаций в области исследования мРНК малым миром.

Поскольку в малом мире вершины достижимы друг до друга за малое количество переходов, такие процессы, как распространение инфекции или знания, должны происходить иначе, чем в обычном графе.

Для определения того, что граф является малым миром, в нескольких работах были предложены различные критерии (Watts, Strogatz, 1998; Newman et al., 2000). В нашей работе мы выбрали категориальный критерий для выявления присутствия малого мира в сети организаций мРНК, следуя (Humphries, Gurney, 2008), где авторы вводят меру «малости мира»:

$$S = \frac{CC_G}{CC_{\text{rand}}} / \frac{L_G}{L_{\text{rand}}}.$$

В уравнении выше  $CC_G$  – коэффициент кластеризации графа  $G$ ;  $L_G$  – средняя длина кратчайшего пути графа  $G$ ;  $CC_{\text{rand}}$  и  $L_{\text{rand}}$  – параметры случайного графа со случайно равномерным размещением ребер с тем же количеством узлов и ребер, что и граф  $G$ .

Процесс распространения знаний можно интерпретировать как процесс «заражения идеями», при котором через промежуточного хозяина (научные публикации) организации могут вдохновиться какой-либо областью исследований и сами начать публиковать статьи. Такой процесс можно моделировать с помощью модели Susceptible, Infectious, Recovered (SIR) (Goffman, Newill, 1964), в рамках которой составляется система дифференциальных уравнений, моделирующих динамику заражения и выздоровления субъектов. В простейшем случае однородной среды решением этих уравнений на малых временах является экспоненциальный рост числа зараженных субъектов.

В работе (Vazquez, 2006) автор моделирует рост заболеваемости с использованием SIR для задач, где графы передачи известны и обладают свойством малого мира (Muldoon et al., 2016). Автор адаптирует модель распространения SIR к представлению исходного графа в виде остова дерева (AST) и получает точную нормализованную частоту заболеваемости для AST,  $\rho(t)$ , которая аппроксимирует эту частоту для исходного графа. Таким образом, учитывая, что граф обладает свойством малого мира, существует точное решение нормализованной частоты заражения для AST, которое является аппроксимацией для исходного графа:

$$\rho(t) = \lambda \frac{(\lambda t)^{D-1}}{(D-1)!} e^{-(\lambda+\mu)t} \left[ 1 + O\left(\frac{t_0}{t}\right) \right],$$

где  $\lambda$  и  $\mu$  – соответственно скорости заражения и выздоровления в рамках SIR модели;  $D$  – среднее кратчайшее расстояние графа;  $t_0$  – время перехода между режимами. Для этого граф, помимо характеристики малого мира, должен удовлетворять одному из условий для  $\gamma$  (показатель степенного закона распределения степеней вершин) и  $\nu$  (коэффициент корреляции Пирсона степени между парами соединенных узлов) (Vazquez, 2006):

$$\begin{aligned} \gamma &> 3, \quad \nu > 0, \\ 2 \leq \gamma \leq 3, \quad \nu > -1, \quad 3 - \gamma + \nu > 0. \end{aligned}$$

## Материалы и методы

Для сбора набора данных аффилиаций области исследования мРНК использовалась цифровая библиотека PubMed. Из аффилиаций были выделены упоминания организаций. При этом использовался подход, основан-

ный на ключевых словах, для понимания, какая часть аффилиации содержит какую информацию об упоминании организации (название организации, страна, город и т. д.):

Пример разбиения аффилиации на упоминания организаций с определением страны для статьи с PubMed ID 19996210

- |   |  |
|---|--|
| <p>(1) Authors' Affiliations: Cancer Genetics, Kolling Institute of Medical Research; Department of Endocrinology; Department of Anatomical Pathology, Royal North Shore Hospital, St. Leonards, New South Wales, Australia; Department of Surgery, Bankstown Hospital, Bankstown, New South Wales, Australia; South Western Sydney Clinical School, University of New South Wales; Endocrine Surgical Unit, University of Sydney; Department of Surgery, Liverpool Hospital, Sydney, New South Wales, Australia; Endocrine Surgical Unit, University of California Los Angeles; and Division of Hematology and Oncology, Department of Medicine, University of California Los Angeles School of Medicine, Los Angeles, California.</p> | <ol style="list-style-type: none"> <li>1. kolling institute of medical research, Australia</li> <li>2. royal north shore hospital, Australia</li> <li>3. bankstown hospital, Australia</li> <li>4. university of new south wales, Australia</li> <li>5. university of sydney, Australia</li> <li>6. liverpool hospital, Australia</li> <li>7. university of california los angeles, UNKNOWN</li> <li>8. university of california los angeles, school of medicine, UNKNOWN</li> </ol> |
|---|--|

Затем для всех упоминаний был построен словарь уникальных К-Мер (n-gram), где  $K = 2$ , и для каждого упоминания сформирован булевый вектор присутствия определенного К-Мер в этом упоминании. Далее эти векторы упоминаний были отсортированы лексикографически,

чтобы получить список векторов, в котором схожие упоминания сгруппированы по построению. После чего для каждой соседней пары упоминаний было посчитано расстояние по метрике Dice. Если оно превышало заданный порог, это было свидетельством того, что упоминания относятся к различным кластерам, что дает нам группировку упоминаний (см. таблицу).

Сгруппированные упоминания содержат ссылки на одну и ту же организацию, поэтому на следующем шаге мы можем построить граф соавторства организаций, определяя, какие организации опубликовали одну и ту же статью вместе.

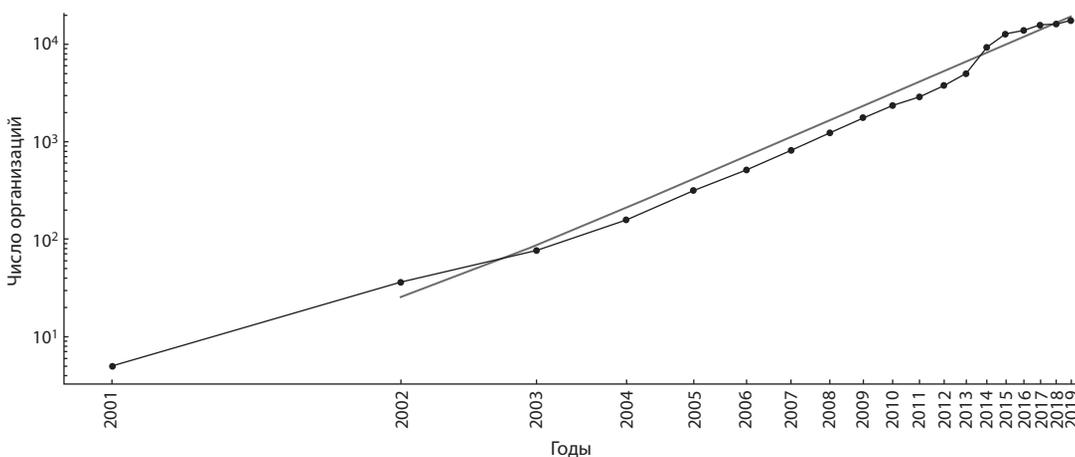
### Результаты

Анализ структурных характеристик графа научных организаций в области исследования микроРНК показывает, что этот граф удовлетворяет критериям малого мира (Muldoon et al., 2016) с показателем степени степенного распределения  $\gamma = 2.01$  и коэффициентом ассортативности связности вершин графа  $\nu = -0.03$ . Поэтому для числа научных организаций, имеющих публикации в данной области, можно ожидать степенной рост согласно модели (Vazquez, 2006). Из (Vazquez, 2006) следует, что начальный рост числа вершин имеет степенную зависимость с показателем степени  $D - 1$ , где  $D$  – средняя длина кратчайшего пути в графе. Для графа научных организаций области исследования микроРНК  $D = 3.46$ , а аппроксимированный степенной параметр  $D - 1 = 2.64 \pm 0.23$  (см. рисунок), что

Пример идентификации организаций

#	Упоминание	2-Мер булевый вектор	Метрика Dice
1	institute	1111111100000000	0.2
2	insitute	1111100100001000	0.429
3	institue	1111011000010000	<b>0.834</b>
4	center	0000100011100100	0.4
5	centre	000000011100011	

Примечание. Значение порога – 0.8,  $K = 2$ . Расстояние между элементами 3, 4 превышает значение порога, что приводит к разбиению элементов на кластеры. Примеры 2-Мер – in, ns, st, ti, it, tu, ...



Ежегодное количество организаций, опубликовавших статью в области исследования микроРНК в зависимости от времени в двойных логарифмических координатах.

для показателя степени дает отклонение около 7 % от предсказанного моделью.

При аппроксимации скорости «заражения информацией» получена величина скорости  $\lambda = 0.184 \pm 0.002 \text{ год}^{-1}$ , которая характеризует скорость публикации первой статьи по мРНК у организации при ее взаимодействии с другой организацией, уже публиковавшейся в этой области.

Анализ подграфа российских научных институтов в области исследования мРНК показывает, что активность российских организаций уступает средней активности организаций (среднее количество публикаций на одну организацию в России составляет 0.92 против среднего по области значения 21.5). При этом российское сообщество оказалось более плотным: коэффициент кластеризации подграфа российских организаций превышает средний по области: 0.708 в России и 0.361 в среднем. Самый активный партнер России в международном сотрудничестве – США, с 50 совместными публикациями. Однако американо-российское сотрудничество нестабильно и децентрализовано, а лидерами по активности сотрудничества с российскими организациями являются Немецкий центр изучения рака, Харбинский медицинский университет и Каролинский институт (по 6 совместных публикаций).

## Обсуждение

Понимание факторов продуктивности исследовательских организаций и динамики их публикационной активности важно для управления наукой. Помимо алгоритмов автоматической идентификации организаций, активно развиваются такие проекты, как [gog.org](http://gog.org), направленные на идентификацию научных институтов за счет присвоения им уникальных идентификаторов (подобно [orcid.org](http://orcid.org) для авторов). Данные проекты упрощают поиск организаций, но требуют принятия использования таких проектов авторами публикаций, так как для возможности полной идентификации каждой организации необходимо указывать идентификатор [gog.org](http://gog.org) для каждой аффилиации из публикации, что на текущий момент не может быть гарантировано. Поэтому в ближайшее время алгоритмы автоматической идентификации организаций останутся актуальными.

В работе данные представлены по 2019 г. К настоящему времени структура графа могла измениться. Кроме того, информация о статьях в библиотеке PubMed может обновляться ретроспективно. Тем не менее данные по публикациям на 23.01.2022 показывают, что картина эволюции области мРНК принципиально не изменилась (данные не приведены). Новая геополитическая реальность неизбежно скажется на структуре взаимодействия и соавторства в научных областях. Однако в связи с запаздыванием по времени видимых результатов изменение

в научном сотрудничестве проявится в базах данных не ранее 2024 г.

## Заключение

Одной из моделей развития новых областей знания является модель «интеллектуальной эпидемии», в которой новые идеи случайно распространяются среди исследователей, заражая все большее и большее их число (Goffman, Newill, 1964). Закон распространения может определяться структурой среды. В нашей работе показано, что граф взаимодействия организаций области исследования мРНК является малым миром, и вследствие этого публикационная активность области демонстрирует степенной рост согласно модели (Vazquez, 2006). Более медленный по сравнению с экспоненциальным, степенной рост возникает из-за «самоизбегания» путей распространения в компактных сетях малого мира: при «заражении информацией» очередного узла малого мира высока вероятность того, что этот узел уже был «заражен» альтернативным путем. Граф соавторства для нашего анализа был построен с использованием алгоритма кластеризации упоминаний организаций на основе сортировки булевых векторов признаков К-Мер.

## Список литературы / References

- Goffman W., Newill V.A. Generalization of epidemic theory. An application to the transmission of ideas. *Nature*. 1964;204(4955):225-228. DOI 10.1038/204225a0.
- Humphries M.D., Gurney K. Network 'small-world-ness': a quantitative method for determining canonical network equivalence. *PLoS One*. 2008;3(4):e0002051. DOI 10.1371/journal.pone.0002051.
- Leydesdorff L., Wagner C., Park H., Adams J. International collaboration in science: the global map and the network. *Prof. Inf.* 2013; 22(1):1-18. DOI 10.3145/epi.2013.ene.12.
- Liu M., Li D., Qin P., Liu C., Wang H., Wang F. Epidemics in interconnected small-world networks. *PLoS One*. 2015;10(3):e0120701. DOI 10.1371/journal.pone.0120701.
- Muldoon S., Bridgeford E., Bassett D. Small-world propensity and weighted brain networks. *Sci. Rep.* 2016;6:22057. DOI 10.1038/srep22057.
- Newman M.E.J., Moore C., Watts D.J. Mean-field solution of the small-world network model. *Phys. Rev. Lett.* 2000;84(14):3201-3204. DOI 10.1103/PhysRevLett.84.3201.
- Ribeiro L., Rapini M., Silva L., Albuquerque E.M. Growth patterns of the network of international collaboration in science. *Scientometrics*. 2018;114:159-179. DOI 10.1007/s11192-017-2573-x.
- Shi Y., Guan J. Small-world network effects on innovation: evidences from nanotechnology patenting. *J. Nanopart. Res.* 2016;18:329. DOI 10.1007/s11051-016-3637-1.
- Vazquez A. Spreading dynamics on small-world networks with connectivity fluctuations and correlations. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* 2006;74:056101. DOI 10.1103/PhysRevE.74.056101.
- Wagner C., Leydesdorff L. Network structure, self-organization and the growth of international collaboration in science. *Res. Policy*. 2005; 34(10):1608-1618. DOI 10.1016/j.respol.2005.08.002.
- Watts D.J., Strogatz S.H. Collective dynamics of 'small-world' networks. *Nature*. 1998;393(6684):440-442. DOI 10.1038/30918.

## ORCID ID

A. Firsov [orcid.org/0000-0002-7681-1032](https://orcid.org/0000-0002-7681-1032)  
I.I. Titov [orcid.org/0000-0002-2691-3292](https://orcid.org/0000-0002-2691-3292)

**Благодарности.** Работа ИТ была поддержана в рамках государственного бюджетного проекта РФ ФВНР-2022-0020.

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 07.09.2022. После доработки 10.11.2022. Принята к публикации 10.11.2022.

# Алфавитный указатель авторов статей, опубликованных в журнале в 2022 г.

- Агеева Е.В. 7, 675  
Адамовская А.В. 8, 733  
Айдаров А.Н. 5, 413  
Акинъшин А.А. 8, 758  
Аклеев А.В. 1, 50  
АлБосале А.Х. 1, 59  
Алексеев Я.И. 6, 544  
Аль-Накиб Е.А. 7, 645  
Амстиславская Т.Г. 4, 365  
Амстиславский С.Я. 4, 365; 5, 431  
Андреева Л.В. 6, 537  
Анисимова Е.А. 6, 560  
Антонов А.А. 4, 349  
Антонова О.Ю. 2, 128  
Антропова Е.А. 8, 733  
Ануфриев К.Э. 8, 743  
Атопкин Д.М. 3, 290  
Аульченко Ю.С. 4, 378  
Афанасенко О.С. 3, 272  
Афонников Д.А. 8, 787  
Афонникова С.Д. 8, 810  
Аюпова Н.Б. 8, 758
- Багиров В.А. 4, 378  
Бажан Н.М. 2, 159  
Бакулина А.Ю. 3, 240  
Балабова Д.В. 3, 240  
Балыбина Н.Ю. 2, 159  
Баранов К.О. 2, 159  
Баттулин Н.Р. 4, 402  
Белавин П.А. 3, 327  
Белан И.А. 7, 662  
Беленькая С.В. 3, 240  
Белкасем эль-Амрани 5, 442  
Белов А.Н. 3, 240  
Беловежец Л.А. 6, 568  
Белчев И. 2, 139  
Белькова Н.Л. 5, 495  
Берман Д.И. 1, 109  
Блинова Е.А. 1, 50  
Блинова С.А. 6, 544  
Богданова В.С. 4, 359  
Богомоллов А.Г. 1, 96; 8, 798  
Богословская Т.Ю. 3, 319  
Бойко А.П. 7, 599  
Бондарь А.А. 3, 240  
Бородин П.М. 4, 378  
Бочарникова М.Е. 8, 787  
Брусенцев Е.Ю. 4, 365; 5, 431  
Брызгалов Л.О. 1, 65  
Букин Ю.С. 1, 74, 86  
Булахова Н.А. 1, 109  
Бурменко Ю.В. 7, 622  
Бурыгин Г.Л. 5, 422  
Быков Р.А. 6, 553
- Васильев В.Б. 3, 319  
Васильева О.Ю. 1, 22  
Вензель А.С. 8, 733  
Вергунов Е.Г. 8, 765  
Весна Э. 8, 806  
Вишневецкий О.В. 8, 798, 819  
Вишнякова М.А. 7, 599  
Водясова Е.А. 3, 290  
Волкова Н.А. 4, 378  
Воловик В.Т. 4, 349  
Воробьева С.С. 5, 467  
Ворожейкин П.С. 8, 819  
Воронина А.В. 2, 146  
Воронина Е.Н. 2, 188  
Воропаева Н.М. 5, 495
- Гавриленко А.Д. 8, 733  
Гаврилова Е.В. 4, 394  
Газизова Г.Р. 1, 109  
Герасимова С.В. 2, 153  
Герасимчук А.Л. 5, 449  
Гисматулина Ю.А. 1, 22  
Глушаков Д.А. 7, 609  
Голохваст К.С. 7, 637  
Голубев С.Н. 5, 477  
Голубятников В.П. 8, 758  
Гончаров Н.П. 7, 662  
Горбунова М.Е. 6, 560  
Григорова Е.В. 5, 495  
Грин И.Р. 4, 341  
Гульятеева Е.И. 6, 537  
Гурина В.В. 6, 568
- Давыдова Ю.Д. 2, 179  
Дамаров И.С. 1, 65  
Деятериков А.П. 8, 780  
Дейнеко Е.В. 3, 327  
Деменков П.С. 8, 733  
Джос Е.А. 7, 652  
Дмитриева Е.В. 3, 290  
Додонова Е.А. 6, 560  
Дорогина О.В. 1, 22  
Драгов Р.Г. 6, 515  
Драчкова И.А. 3, 227  
Дружин А.Е. 6, 537  
Дубинина А.Д. 2, 159  
Душкин В.А. 4, 349  
Дьяченко Е.А. 7, 652  
Дюдеева Е.С. 1, 5
- Евдокименко С.Н. 7, 622  
Евдокимов М.Г. 7, 609  
Евсеева Н.В. 5, 422  
Егорова А.А. 2, 153  
Елаткин Н.П. 3, 298

Елизарова И.А. 6, 560  
Елизарова С.А. 6, 583  
Ельчанинов В.В. 3, 240  
Еникеева Р.Ф. 2, 179  
Ермаков М.С. 1, 14  
Ефимов В.М. 7, 662  
Ефремов Г.И. 6, 507

**Жарков Д.О.** 4, 341  
Жданова И.Н. 5, 486  
Жемчужина Н.С. 6, 583

Зайнуллин Л.И. 6, 560  
Захаренко А.М. 7, 637  
Захарова Ф.М. 3, 319  
Зеленских М.О. 8, 773  
Землянская Е.В. 8, 721, 798  
Злобин А.С. 4, 378  
Золотарева К.А. 8, 798  
Зуев Е.В. 7, 662  
Зуева Г.А. 1, 22

**Иванисенко В.А.** 2, 121; 8, 733  
Иванисенко Н.В. 8, 733  
Иванисенко Т.В. 8, 733  
Ивасенко Д.А. 5, 449  
Игонина Т.Н. 5, 431  
Игошин А.В. 3, 298  
Илинский Ю.Ю. 6, 553  
Ильичев А.А. 2, 214  
Иолчиев Б.С. 4, 378

**Кабиллов М.Р.** 1, 86  
Казанцева А.В. 2, 179  
Казанцева А.Ю. 2, 159  
Казачек А.В. 8, 798  
Капустина И.С. 6, 568  
Каргаполова К.Ю. 5, 422  
Карлов А.В. 7, 704  
Карпенко Л.И. 2, 214  
Карссен Л.С. 4, 378  
Касьмова А.А. 5, 449  
Керв Ю.А. 7, 630  
Кириченко А.В. 4, 378  
Кирьякова М.Н. 7, 609  
Киселева Е.В. 2, 169  
Киселева М.И. 6, 583  
Клименко И.А. 4, 349  
Ковалева О.Н. 6, 524  
Коваль А.Д. 3, 240  
Коваль О.А. 1, 14  
Ковас Ю.В. 2, 179  
Коломиец Т.М. 6, 583  
Колосова И.В. 4, 394  
Колчанов Н.А. 8, 719, 743  
Комиссаров А.С. 8, 810  
Конарев А.В. 7, 630  
Конькова Н.Г. 7, 637  
Корболина Е.Е. 1, 65  
Королева М.Л. 6, 544

Костерин О.Э. 4, 359  
Котикова А.И. 1, 50  
Кочетов А.В. 2, 153; 3, 250  
Кочиева Е.З. 6, 507  
Кравцова Л.С. 1, 86  
Кривина Е.С. 1, 74  
Кручинина Ю.В. 7, 662  
Кузнецова Л.И. 4, 385  
Кузнецова М.В. 5, 486  
Кузьмина Т.И. 3, 234  
Кулакова А.В. 6, 507; 7, 652  
Куликов И.М. 7, 622  
Курочкин В.Е. 6, 544  
Кучур П.Д. 8, 810

**Лаврик И.Н.** 8, 733  
Ларичев К.Т. 3, 250  
Ларкин Д.М. 3, 298  
Лашина Н.М. 3, 272  
Лебедкин Д.А. 8, 765, 773  
Левинсон А.Л. 5, 431  
Леонова И.Н. 7, 675  
Лепехов С.Б. 2, 196  
Липина Т.В. 4, 365  
Лобакова Е.С. 6, 575  
Лободина Е.В. 7, 645  
Локачук М.Н. 4, 385  
Лоскутов И.Г. 6, 524; 7, 597  
Лукина К.А. 6, 524  
Льобунь Е.В. 5, 477

**Мавлютов Ю.М.** 4, 349  
Макарова Е.Н. 2, 159  
Максютов Р.А. 4, 394  
Малых С.Б. 2, 179  
Мальшев Л.Л. 7, 630  
Малькеева Д.А. 2, 169  
Мандельштам М.Ю. 3, 319  
Манучарова Н.А. 6, 575  
Маркова Ю.А. 6, 568  
Матвеева Т.В. 7, 697  
Матора Л.Ю. 5, 422  
Матушкин Ю.Г. 1, 96; 8, 719  
Машкина Е.В. 1, 59  
Мглинец А.В. 4, 359  
Меркулова Т.И. 1, 65  
Мешкова Л.В. 7, 609  
Мещерякова Е.Н. 1, 109  
Минушкина Л.С. 8, 758  
Минчева Е.В. 1, 86  
Миргазов Д.А. 6, 560  
Мироненко Н.В. 3, 272  
Митрофанова О.П. 2, 128  
Михайловская В.С. 5, 486  
Мищенко Е.Л. 2, 121  
Моргунов А.И. 5, 413  
Мотылева С.М. 7, 622  
Муратова А.Ю. 5, 477  
Мустафин Р.Н. 1, 40; 2, 179; 3, 308

- Назарова Г.Г. 4, 371  
Нгуен М.Л. 2, 146  
Немцев Б.Ф. 7, 662  
Немченко У.М. 5, 495  
Никифоров В.С. 1, 50  
Никонова А.А. 5, 467  
Новикова Л.Ю. 3, 264; 7, 599  
Нуждина Н.С. 1, 22  
Нурминский В.Н. 6, 568  
Нуштаева А.А. 1, 14
- Овчинникова Е.С. 7, 662  
Озерский П.В. 3, 264  
Озолина Н.В. 6, 568  
Окотруб С.В. 4, 365  
Омельянчук Н.А. 8, 721  
Осадчук А.В. 1, 96  
Осадчук Л.В. 1, 96  
Осипова Л.П. 2, 188  
Осянин К.А. 6, 560  
Ощепков Д.Ю. 1, 96
- Павловская Е.Н. 4, 385  
Пальянов А.Ю. 8, 780  
Панищева Д.В. 7, 622  
Панкратова Л.Ф. 6, 583  
Парахина О.И. 4, 385  
Пахотина И.В. 7, 609  
Пельтек С.Е. 1, 109  
Перетолчина Т.Е. 1, 86  
Пермякова Н.В. 4, 341  
Петрова Д.В. 4, 341  
Пискарев В.В. 7, 662  
Плеканчук В.С. 3, 281  
Подгаецкий М.А. 7, 622  
Подколodная О.А. 8, 798  
Подколodный Н.Л. 8, 798  
Полубоярова Т.В. 1, 109  
Пономаренко М.П. 1, 96; 3, 227; 8, 798  
Пономаренко П.М. 1, 96; 8, 798  
Попов В.С. 7, 630  
Поротников И.В. 2, 128  
Потоцкая И.В. 5, 413  
Притужалова А.О. 3, 234  
Прокудина О.И. 3, 281  
Проскурняк Л.П. 4, 371  
Пышная И.А. 1, 5  
Пышная О.Н. 7, 652
- Разгонова М.П. 7, 637  
Ракевич А.Л. 6, 568  
Рассказов Д.А. 8, 798  
Ревва П.М. 8, 733  
Ремезовская Н.Б. 5, 486  
Рихтер В.А. 1, 14  
Рожкова И.Н. 4, 365; 5, 431  
Ромашов Г.А. 3, 298  
Рудометов А.П. 2, 214; 3, 240  
Рудометова Н.Б. 2, 214  
Рудыч П.Д. 8, 773
- Рухлова Е.А. 3, 240  
Рябинин А.С. 6, 553  
Рязанова М.А. 3, 281
- Савилов Е.Д. 5, 495  
Савинкова Л.К. 3, 227; 8, 798  
Савкина О.А. 4, 385  
Савостьянов А.Н. 8, 765, 773  
Салина Е.А. 3, 250  
Сапрыгин А.Е. 8, 765, 773  
Свистунова Н.Ю. 7, 622  
Семенова Е.В. 7, 599  
Сергеева Е.М. 3, 250  
Серов О.Л. 4, 402  
Сибикеев С.Н. 6, 537  
Сидоренко А.Д. 8, 721  
Сидоров А.В. 6, 568  
Ситникова К.О. 5, 495  
Смирнов Д.Н. 1, 109  
Смоленская А.Е. 7, 630  
Смоленская С.Э. 7, 662  
Соколова Д.В. 1, 30  
Старич Эрьявец М. 5, 486  
Стоянов Х. 2, 139  
Супрун И.И. 7, 645  
Суслов В.В. 1, 96  
Сухарева Е.С. 5, 495  
Сухорева М.В. 5, 495
- Табиханова Л.Э. 2, 188  
Таможников С.С. 8, 773  
Татарская Д.Н. 3, 234  
Тахирова З.Р. 2, 179  
Темралеева А.Д. 1, 74  
Титов И.И. 8, 819, 826  
Тихонова Н.Г. 7, 630  
Тихонова О.А. 7, 630  
Ткаченко А.Г. 5, 458  
Ткаченко О.В. 5, 422  
Токмаков С.В. 7, 645  
Трегубчак Т.В. 4, 394  
Тумаева Т.А. 7, 622  
Тупикин А.Е. 1, 86  
Турковская О.В. 5, 477  
Турнаев И.И. 8, 787
- Уварова Е.А. 3, 327  
Ульданова Е.Е. 4, 365  
Ульянов А.В. 7, 704
- Фахрутдинов Н.А. 6, 560  
Федотов А.П. 1, 86  
Фёдорова С.А. 2, 169  
Филипенко М.Л. 2, 188  
Фирсов А.Б. 8, 826  
Фишман В.С. 8, 806  
Франк Ю.А. 5, 449
- Хаммадов Н.И. 6, 560  
Хандаев Б.М. 8, 798

Хатевфов Э.Б. 7, 704  
Хафизова Г.В. 7, 697  
Хлебодарова Т.М. 8, 733, 743  
Хлесткин В.К. 4, 385  
Хлесткина Е.К. 7, 597  
Храмеева Е.Е. 1, 109  
Хуен Т.Н.Б.Т. 2, 146  
Хуснутдинова Э.К. 1, 40; 2, 179  
Хютти А.В. 3, 272

**Ц**епилов Я.А. 4, 378  
Цыганов И.В. 5, 458

**Ч**адаева И.В. 1, 96; 3, 227; 8, 798  
Челебиева Э.С. 3, 290  
Чепурнов Г.Ю. 7, 662  
Черняева Е.Н. 3, 298  
Чешкова А.Ф. 2, 202  
Чжоу Чэньси 7, 662  
Чинь Д.М. 2, 146  
Чистякова И.В. 3, 234  
Чуйко Э.А. 4, 365  
Чуркина Т.В. 2, 188

**Ш**агимарданова Е.И. 1, 109  
Шаманин В.П. 5, 413  
Шамустакимова А.О. 4, 349

Шапиро Т.Н. 6, 575  
Шарыпова Е.Б. 1, 96; 3, 227; 8, 798  
Шашкова Т.И. 4, 378  
Швалов А.Н. 4, 394  
Шварцев А.А. 6, 544  
Шеленга Т.В. 7, 630  
Шереметьева М.Е. 8, 743  
Шеховцов С.В. 1, 109  
Шипова А.А. 1, 109  
Широков А.А. 5, 422  
Шихат О.В. 3, 290  
Шишкина О.Д. 6, 553

**Щ**елкунов С.Н. 4, 394  
Щенникова А.В. 6, 507; 7, 652  
Щербаков Д.Н. 2, 214; 3, 240  
Щербаков Д.Ю. 1, 86  
Щербань А.Б. 7, 684  
Щёголев С.Ю. 5, 422

**Ю**дин Н.С. 3, 298  
Юсов В.С. 7, 609

**Я**ковлева Т.В. 2, 159  
Якубицкий С.Н. 4, 394  
Яненко А.С. 8, 743

Прием статей через электронную редакцию на сайте <http://vavilov.elpub.ru/index.php/jour>  
Предварительно нужно зарегистрироваться как автору, затем в правом верхнем углу страницы выбрать «Отправить рукопись». После завершения загрузки материалов обязательно выбрать опцию «Отправить письмо», в этом случае редакция автоматически будет уведомлена о получении новой рукописи.

«Вавиловский журнал генетики и селекции»/“Vavilov Journal of Genetics and Breeding”  
до 2011 г. выходил под названием «Информационный вестник ВОГиС»/  
“The Herald of Vavilov Society for Geneticists and Breeding Scientists”.

Регистрационное свидетельство ПИ № ФС77-45870 выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 20 июля 2011 г.

«Вавиловский журнал генетики и селекции» включен ВАК Минобрнауки России в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук, Российский индекс научного цитирования, ВИНТИ, базы данных Emerging Sources Citation Index (Web of Science), Zoological Record (Web of Science), Scopus, PubMed Central, Ebsco, DOAJ, Ulrich’s Periodicals Directory, Google Scholar, Russian Science Citation Index на платформе Web of Science, каталог научных ресурсов открытого доступа ROAD.

Открытый доступ к полным текстам:  
на сайте ИЦиГ СО РАН – <https://sites.icgbio.ru/vogis/>  
платформе Elpub – [vavilov.elpub.ru/index.php/jour](http://vavilov.elpub.ru/index.php/jour)  
платформе Научной электронной библиотеки – [elibrary.ru/title\\_about.asp?id=32440](http://elibrary.ru/title_about.asp?id=32440)  
PubMed Central, <https://www.ncbi.nlm.nih.gov/pmc/journals/3805/>

Подписку на «Вавиловский журнал генетики и селекции» можно оформить в любом почтовом отделении России. Индекс издания 42153 по каталогу «Пресса России».

При перепечатке материалов ссылка на журнал обязательна.

✉ e-mail: [vavilov\\_journal@bionet.nsc.ru](mailto:vavilov_journal@bionet.nsc.ru)

Издатель: Федеральное государственное бюджетное научное учреждение  
«Федеральный исследовательский центр Институт цитологии и генетики  
Сибирского отделения Российской академии наук»,  
проспект Академика Лаврентьева, 10, Новосибирск, 630090.

Адрес редакции: проспект Академика Лаврентьева, 10, Новосибирск, 630090.

Секретарь по организационным вопросам С.В. Зубова. Тел.: (383)3634977.

Издание подготовлено информационно-издательским отделом ИЦиГ СО РАН. Тел.: (383)3634963\*5218.

Начальник отдела: Т.Ф. Чалкова. Редакторы: В.Д. Ахметова, И.Ю. Ануфриева. Дизайн: А.В. Харкевич.

Компьютерная графика и верстка: Т.Б. Коняхина, О.Н. Савватеева.

Подписано в печать 20.12.2022. Выход в свет 30.12.2022. Формат 60 × 84 1/8. Усл. печ. л. 13.95.

Уч.-изд. л. 15.5. Тираж 150 экз. (1-й завод 1–45 экз.) Заказ № 360. Цена свободная.

Отпечатано в Сибирском отделении РАН, Морской проспект, 2, Новосибирск, 630090.



