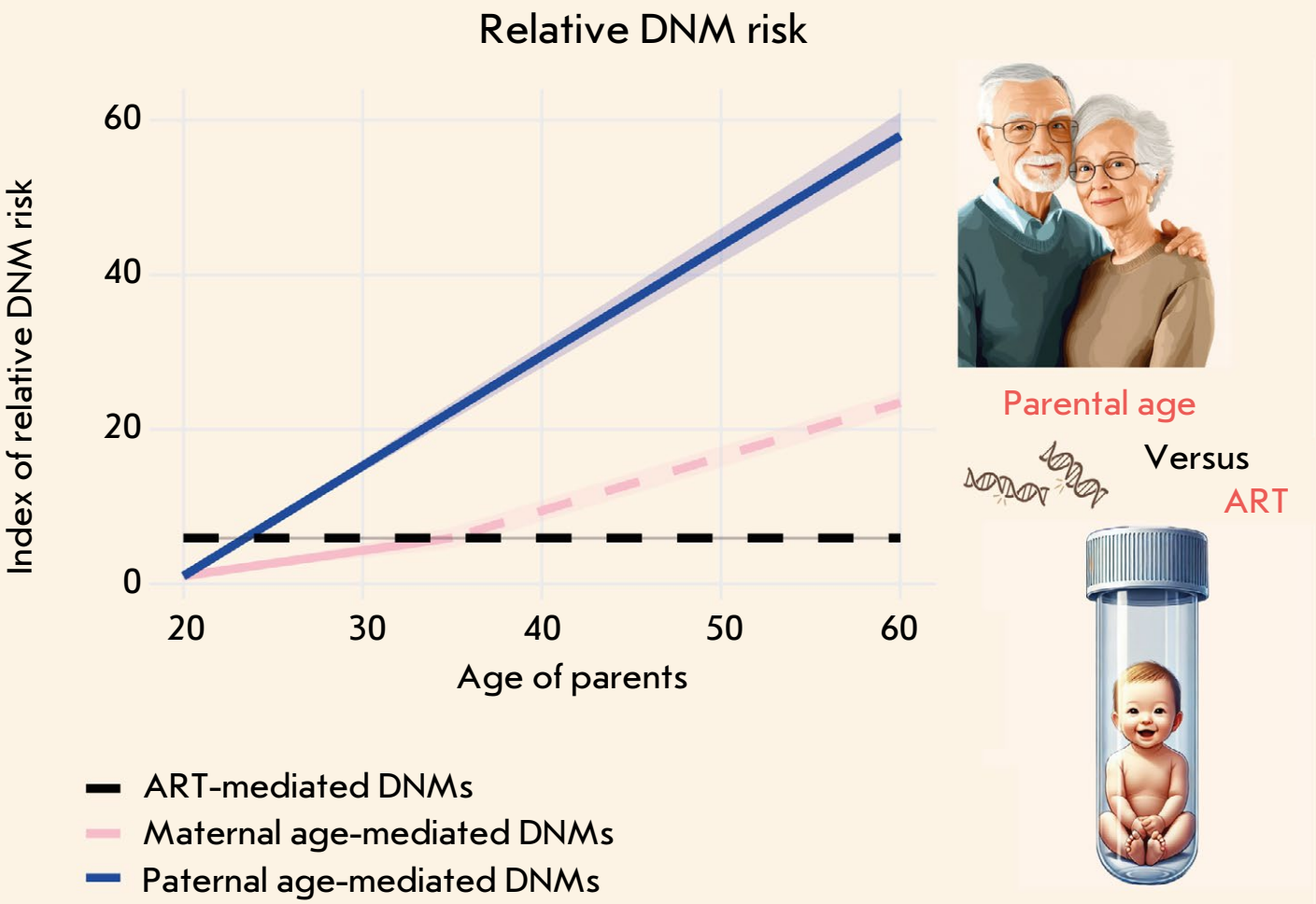
APRIL–JUNE 2025 VOL. 17 № 2 (65)

ISSN 2075-8251

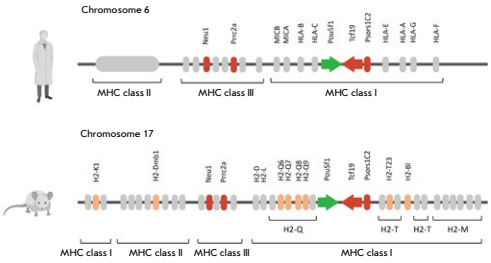
ActaNaturae

The Effects of Assisted Reproductive Technologies on *De Novo* Mutations



Cis-regulatory Function of the *Pou5f1* Gene Promoter in the Mouse MHC Locus

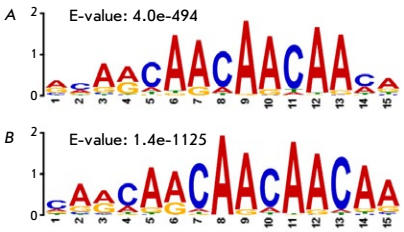
V. V. Ermakova, E. V. Aleksandrova, A. A. Kuzmin, A. N. Tomilin



A schematic depiction of the *Pou5f1-MHC* locus for human and mouse

The *Pou5f1* gene encodes the Oct4 protein, one of the key transcription factors required for maintaining the pluripotent state of epiblast cells and the viability of germ cells. In this study, the authors developed a genetic model based on mouse embryonic stem cells (ESCs) to assess the role of the *Pou5f1* gene promoter in the transcriptional regulation of neighboring genes within the major histocompatibility complex (MHC) locus. They demonstrated that deletion of this promoter does not affect the expression of the selected genes in this locus, either in ESCs or in their trophoblast derivatives. A notable exception is the *Tcf19* gene, which is upregulated upon *Pou5f1* promoter deletion and may be associated with the atherosclerosis pathology due to its pro-inflammatory activity. The developed genetic model will pave the way for future studies into the functional contribution of the *cis*-regulatory association of *Pou5f1, Tcf19*, and, possibly other genes with the atherosclerotic phenotype previously reported in mice carrying the *Pou5f1* promoter deletion in vascular endothelial and smooth muscle cells.

The Drosophila Zinc Finger Protein Aef1 Colocalizes with Enhancers and Is Involved in the Transcriptional Regulation of Numerous Genes

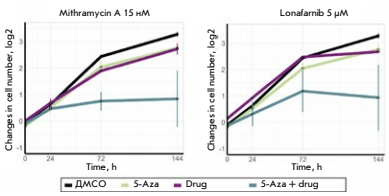


The potential consensus binding motif of the Aef1 protein identified at promoter sites (*A*) and outside promoters (*B*)

N. E. Vorobyeva, J. V. Nikolenko, A. N. Krasnov

The authors found that Aef1 binding sites in Drosophila S2 cells, located outside gene promoters, are nucleosome-depleted regions and colocalize with the SAGA, dSWI/SNF, and ORC complexes. Aef1 binding sites colocalize with the CBP protein and the H3K27Ac histone tag, which is considered to be an active enhancer mark. An RNA-Seq experiment was conducted in Drosophila S2 cells, both normal and with RNA interference targeting the Aef1 protein, to study the role played by the Aef1 protein in transcriptional regulation. The Aef1 protein was shown to affect the transcription of 342 genes, more than half of those (178 genes) containing Aef1 at their promoters or enhancers. Hence, the authors infer that the Aef1 protein is recruited to both promoters and enhancers and is involved, both directly and indirectly, in the regulation of the transcription of the respective genes.

The Hypomethylating Agent 5-Azacitidine Potentiates the Effect of RAS and Sp1 Inhibitors in Neuroblastoma Cells



The effectiveness of combinations of 5-Aza and antitumor drugs against human neuroblastoma SH-SY5Y cells

K. A. Ivanenko, A. V. Snezhkina, M. A. Zolotovskaia, P. V. Spirin, O. G. Leonova, V. I. Popenko, A. V. Kudryavtseva, A. A. Buzdin, V. S. Prassolov, T. D. Lebedev

Neuroblastoma is a malignant, solid tumor caused by the transformation of neural crest cells. DNA hypermethylation of neuroblastoma cells indicates that it is possible to use hypomethylating agents in a combination therapy of the disease. Analysis of transcriptome changes induced by 5-azacitidine (5-Aza) helped identify drugs that exert a synergistic effect on neuroblastoma cell death, particularly the synergistic action of 5-Aza in combination with mithramycin A and lonafarnib in SH-SY5Y neuroblastoma cells. Further research into the efficacy of these drug combinations could focus on a more detailed study of their mechanism of joint action and testing on other neuroblastoma models.

ActaNaturae
APRIL—JUNE 2025 VOL. 17 № 2 (65)
since april 2009, 4 times a year

Founders
Acta Naturae, Ltd,
National Research University
Higher School of Economics

Adress: 101000, Moscow, Myasnitskaya Ulitsa, 13, str. 4
Phone/Fax: +7 (495) 727 38 60
E-mail: actanaturae@gmail.com

Номер подписан в печать 30 июня 2025 г.
Тираж 15 экз. Цена свободная.
Отпечатано в типографии: НИУ ВШЭ,
г. Москва, Измайловское шоссе, 44, стр. 2

Founder and Chairman
of the Editorial Board (from 2009 to 2023)
of the journal Acta Naturae
Academician Grigoriev Anatoly Ivanovich

Indexed in PubMed, Web of Science, Scopus, and RISC

Impact Factor: 2.0 (WOS); 3.5 (Scopus)

CONTENTS

REVIEWS

RESEARCH ARTICLES

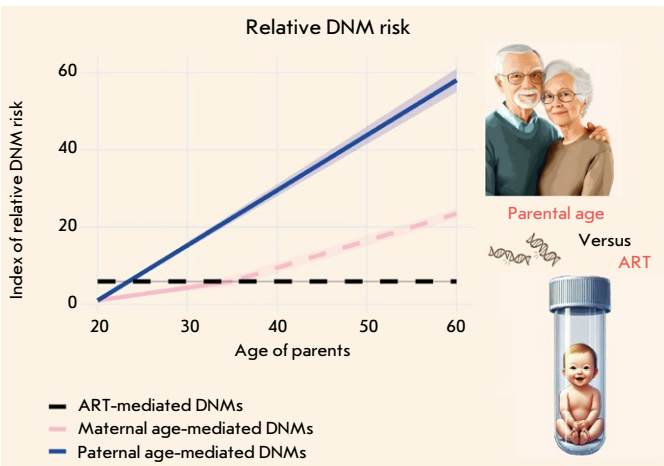


IMAGE ON THE COVER PAGE
(see the article by Arakelyan et al.)

The Effects of Assisted Reproductive Technologies on *De Novo* Mutations

N. A. Arakelyan[1*†], J. Vasilevska[1†], E. I. Rogaev[1,2**]
[1]Center for Genetics and Life Science, Sirius University of Science and Technology, Sirius, 354340 Russia
[2]Department of Psychiatry, UMass Chan Medical School, Worcester, MA 01545 USA
[†]Equal contribution
[*]E-mail: arakelyan.na@talantiuspeh.ru, [**]E-mail: evivrecc@gmail.com
Received: December 11, 2024; in final form, February 13, 2025
DOI: 10.32607/actanaturae.27589

ABSTRACT **Recent advances in assisted reproductive technologies (ART) have revolutionized human reproduction, offering hope to millions of couples facing infertility issues. At the same time, concerns persist regarding the potential impact of ART on the genomic integrity of offspring conceived through these techniques. Specifically, questions abound about the effects of these techniques on the incidence of *de novo* mutations (DNMs), which are genetic alterations that arise spontaneously in the germline or during early embryonic development and are implicated in various human diseases. The extent to which ART directly affects the rate of *de novo* mutations has been the subject of ongoing debate. This review explores recent studies that have investigated the relationship between ART and DNMs. It underscores the necessity for further research to clarify the clinical implications and long-term consequences of ART.**
KEYWORDS **assisted reproductive technologies, *de novo* mutations, aging, paternal age effect, maternal age effect.**

INTRODUCTION

Since the birth of the first child conceived through *in vitro* fertilization more than 45 years ago, the technology has become commonplace in reproductive medicine. The prevalence of couples experiencing infertility has been on a consistent upward trend in recent decades, particularly as an increasing number of individuals postpone their intentions to conceive until a later age. A significant number of these couples ultimately turn to artificial reproductive means. Latest estimates indicate that over 10 million infants have been born globally as a result of assisted reproductive technologies (ART), and that approximately 4 million ART cycles are recorded every year [1, 2].

ART refers to fertility treatments that include all interventions related to the *in vitro* handling of both human oocytes and sperm or embryos for reproduction [3]. Gametes and zygotes undergo a range of non-physiological processes and are exposed to culture media during ART treatments. While the majority of infants conceived through ART are born healthy, concerns persist regarding the safety of these technologies and the potential long-term effects of ART on the development, growth, and overall health of this emerging population [4].

De novo mutations (DNMs), which arise spontaneously in the germline or during early embryonic development, have the potential to shape the genetic outcomes of offspring conceived through ART. A comprehensive understanding of the mechanisms underlying the occurrence of *de novo* mutations and their origin is crucial for addressing the safety and efficacy of ART. Such insights facilitate the development of strategies aimed at mitigating risks and improving ART procedures, with the objective to minimize potential adverse effects on the genetic integrity and health of subsequent generations. The investigation into the relationship between ART and DNMs not only deepens our understanding of the underlying molecular mechanisms involved but also highlights the importance of making informed decisions regarding reproductive interventions [5, 6].

This review seeks to synthesize recent evidence concerning the association between ART and DNMs, given the extensive application of ART and its potential implications for genetic integrity. The objective is

to inform clinical practice and future research in the field of reproductive genetics.

ASSISTED REPRODUCTIVE TECHNOLOGIES (ART)

Assisted reproductive technologies comprise a wide range of procedures aimed at addressing infertility in both females and males. They have witnessed a significant increase in utilization over the years. The approach entails a series of complex steps, including ovarian stimulation, oocyte and sperm preparation, gametes and embryos manipulation, cryopreservation, *in vitro* culture, and, ultimately, embryo transfer. The extent of the intervention can vary considerably, ranging from hormonal treatments for patients to more sophisticated *in vitro* procedures. These interventions encompass a spectrum of techniques, beginning with subtle manipulations of gametes, such as *in vitro* fertilization (IVF), and extending to more invasive procedures, including intracytoplasmic sperm injection (ICSI) and testicular sperm extraction, in conjunction with intracytoplasmic sperm injection (TESE-ICSI). Furthermore, these interventions also involve direct alterations to the embryo, as illustrated by procedures such as trophectoderm or blastomere biopsy (BB). These pivotal stages align with essential developmental phases, characterized by substantial epigenetic remodeling of the genome and increased susceptibility to environmental factors. Any disruption during these critical moments has the potential to adversely impact developmental programming [7, 8]. One key factor contributing to the increased risk associated with ICSI is the sperm selection process. In contrast to conventional IVF, where sperm must independently penetrate the protective layers of the egg, ICSI circumvents these natural selection mechanisms by manually selecting and injecting a single sperm directly into the egg. This process may unintentionally introduce sperm that carry DNA anomalies into the fertilization process. Moreover, the ICSI often involves the retrieval of sperm directly from the testes or epididymis, thereby bypassing the natural maturation process that typically occurs during the sperm transit through the male reproductive tract. The likelihood of genetic abnormalities appearing in offspring may increase as a result of this procedure. Furthermore, the invasive nature of ICSI may elevate the risk of causing damage to the oocyte or the meiotic spindle apparatus during the injection process. Such damage could disrupt DNA replication and repair mechanisms, thereby potentially affecting embryo development [9, 10]. Although the majority of children born through ART are healthy, numerous studies have identified associations between ART and various adverse health outcomes [11–16]. These include congenital heart defects [17] and childhood cancers [18]. However, inconsistencies persist in the literature concerning the increased risks of impaired cognitive development, neurodevelopmental disorders [19], metabolic health, and the effect on reproductive fitness associated with ART [20–22]. Despite the existence of these health risks, it remains unclear whether they are attributable to specific ART procedures, such as sperm retrieval, ovarian stimulation, methods of conception, or laboratory conditions, or whether they are influenced by underlying parental factors or exacerbated by the circumvention of natural selection. Moreover, the interplay between epigenetic and genetic factors in relation to these health outcomes is not yet fully understood. While research on epigenetics within the context of ART is expanding, the investigation of DNMs remains in the nascent stages of exploration. Despite considerable advancements in our understanding of epigenetics, research continues to comprehensively elucidate the role of *de novo* mutations in the context of ART-conceived offspring [23–26].

DE NOVO MUTATIONS

Human *de novo* mutations (DNMs) are defined as the development of new genetic alterations that appear in the germline during one generation. Germline *de novo* mutations (gDNMs) occur in the gametes of one parent and are transmitted directly to the offspring at the time of conception. In contrast, post-zygotic *de novo* mutations occur during the early embryonic development and can affect a significant proportion of the offspring's cells. The implications of a new genetic mutation can vary considerably. While neutral or beneficial mutations may become integrated into the genome of our species, thereby contributing to human evolution, alterations in critical genetic sequences can also disrupt biological systems, potentially resulting in severe diseases. *De novo* mutations are the most severe type of uncommon genetic variations, often more harmful than inherited variations due to less stringent evolutionary selection [5, 27]. They have been linked to various diseases [28], including neurodevelopmental disorders [29–31], heart diseases [32], early-onset high myopia [33], and may also play a role in cases of rare sporadic malformation syndromes like Schinzel–Giedion, Kabuki, and Bohring–Opitz syndromes [34]. Recent studies also indicate their fundamental effects in male and female infertility [35, 36].

De novo mutations can occur at various stages of development, including prior to fertilization in the germ cells, as well as during the cleavage and blastocyst stages of early embryonic development. DNMs encompass a range of mutation types, like point mu-

tations or single-nucleotide variants (SNVs), insertions and deletions (indels) of varying sizes, as well as larger copy number variants (CNVs) and structural variations. Typically, DNMs are identified through the analysis of DNA samples obtained from trios consisting of a father, mother, and offspring. On average, a human genome is estimated to contain between 44 and 82 *de novo* single nucleotide variations, with 1 to 2 of these occurring in coding regions [37–39]. Furthermore, the distribution of DNMs is not random across the genome. Instead, specific genomic features—such as the guanine-cytosine content, nucleosome occupancy, repeat content, recombination rate, transcriptional activity, replication timing, and chromatin state---significantly affect the likelihood of mutations occurring in certain regions. Multiple DNMs may occur in close proximity within the same individual's genome, a phenomenon referred to as mutational clusters. In contrast, mutational hotspots are characterized by the occurrence of multiple DNMs at the same location across several individuals. It has been demonstrated that the rate of genetic mutations in the human germline, encompassing the entire genome, is affected by the specific class of genetic variants in question [40]. The prevalence of germline DNMs in humans significantly varies both within and among families, with these variations influenced by sex. The rate of germline DNMs is believed to be potentially modulated by intrinsic, as well as extrinsic, factors. One of the most important findings concerning germline DNMs is the observation that their frequency increases progressively with the age of both parents at the time of conception, particularly with paternal age. Specifically, there is an increase of approximately 1.35 to 1.5 DNMs in the offspring's germline for each additional year of paternal age. In contrast, the increase in DNMs associated with advancing maternal age is considerably less pronounced, with an addition of only 0.24 to 0.42 DNMs for each additional year of maternal age [41–47]. Approximately 2–3% of all DNMs in the offspring occur in proximity to one another (below 20 kb), forming what are referred to as clustered DNMs (cDNMs). These clusters exhibit a greater maternal bias compared to paternal clusters, which can be attributed to the deficient homologous recombination repair of double-strand breaks associated with the DNA repair mechanisms of aging oocytes. This impairment facilitates the emergence of deregulated recombination hotspots, resulting in mutations that occur in closer proximity than would typically be anticipated. It is noteworthy that 58.4% of maternal cDNMs are enriched within specific genomic regions on the chromosomes 8, 9, and 16, although these regions are also characterized by a higher frequency of maternal unclustered DNMs. In contrast, the frequency of paternal cDNMs correlates with chromosome length. This disparity is linked to the distinct mechanisms that underlie the etiology of *de novo* mutations in different sexes [5, 45].

The predominant hypothesis behind the paternal age effect suggests that mutations arise as a result of random errors during genomic replication. This replication predominantly occurs within the male germ line to facilitate the ongoing production of sperm. Spermatogonia, which serve as the stem cells for spermatogenesis, undergo a series of divisions to both self-renew and produce spermatocytes, which subsequently differentiate into sperm. As males advance in age, the frequency of divisions per spermatogonium rises, thereby increasing the likelihood of new mutations arising from replication errors. Consequently, mutations accumulate in spermatogonia as age advances, thereby illustrating the paternal age effect. While this "selfish selection" model provides insight into the origin of dominant developmental disorders like achondroplasia and Apert syndrome, it fails to extend to the majority of DNMs that do not exhibit an apparent selective bias. Alternatively, recent research has proposed a hypothesis that challenges the traditional perspective, which posits that replication errors are the primary source of human germline mutations. This new hypothesis suggests that the consistently low rate of spontaneous mutations attributed to replication errors associated with paternal age may be augmented by mutations resulting from defective DNA repair mechanisms in the male germ line [5, 37, 48, 49].

The mechanisms underlying the maternal age effect cannot be attributed to genome replication, as oocyte production ceases during the intrauterine phase, and genome replication is no longer active. Therefore, the mechanisms underlying age-related DNMs in mothers must fundamentally differ from those associated with paternal aging. Evidence suggests that maternal age at conception affects the mutation frequency, which may be the result of accumulated damage in the oocytes or potentially materialize via postzygotic mutations occurring in the embryo. Following fertilization, the oocyte has a limited timeframe to repair DNA damage introduced by the spermatozoon prior to the activation of the S-phase in the first mitotic division. Faulty or deficient repair during this process may lead to mutations that impact the health of offspring, a phenomenon referred to as the "post-meiotic oocyte collusion hypothesis" [50–52]. Such a mechanism aligns with recent data indicating that (ART) medical procedures may increase the mutational burden carried by offspring conceived through these procedures [52].

De novo mutations associated with ART

Numerous studies have investigated natural conception and the prevalence of DNMs [53–55]. However, there is a paucity of research examining the relationship between conception through ART and DNMs (*Table 1*). Although many studies generally indicate that ART is safe [56–59], recent evidence suggests that ART may be associated with an increased burden of DNMs in children conceived via this method [6, 20, 35, 60, 61].

In their pivotal study, Wang et al. investigated the potential link between ART and the incidence of congenital heart defects (CHDs) in offspring, with a specific emphasis on DNMs as a contributing factor. The researchers conducted whole-genome sequencing on a cohort comprising 1,137 individuals from 160 families who conceived spontaneously and 205 families who conceived through ART. The findings indicated that children conceived via ART exhibited a significantly higher frequency of germline DNMs (gDNMs) compared to those conceived spontaneously, even after adjusting for various confounding factors such as parental age, smoking habits, alcohol consumption, and levels of physical activity. In particular, offspring conceived through ART exhibited a 4.59-fold increase in germline *de novo* mutations (gDNMs), with 3.32 mutations originating from paternal sources and 1.26 from maternal sources [6]. This finding aligns with a study by Wong et al., which reported that the application of ART for conception resulted in an average of 4.25 additional DNMs per genome compared to natural conception [56]. Notably, paternal DNMs in ART-conceived offspring were characterized by specific C>T substitutions at CpG sites, which have been associated with an elevated risk of CAD. Furthermore, the study indicated that a significant majority (87.9%) of these mutations were inherited from the father, while the utilization of both recombinant and follicle-stimulating hormone, along with a high-dose human chorionic gonadotropin trigger, was correlated with an increase in maternal DNMs [6]. A contrasting study did not observe any significant differences when directly comparing the mitochondrial DNA (mtDNA) variants in oocytes obtained through natural means and those retrieved following ovarian stimulation cycles. Nevertheless, this study identified a correlation between the presence of non-synonymous mtDNA heteroplasmic variants in protein-coding regions and ribosomal RNA loci and a reduction in birth weight. These variants were found to be prevalent in children born to older mothers who had undergone ART treatments. It is hypothesized that these variants may induce a mild, yet significant mitochondrial dysfunction, which could contribute to a lower birth weight percentile. This finding provides preliminary evidence of mitochondrial genetic factors that may help explain the disparities in condition between individuals conceived through ART and those conceived naturally [61].

As previously noted, ICSI has generated greater concerns compared to other ART procedures due to its invasive nature and the potential risks associated with the selection and manipulation of sperm. However, it is important to acknowledge that there exists contradictory data on this matter [23, 63, 64]. In this context, several research groups have studied the specific effect of ICSI on DNMs. A small study conducted by Woldringh et al. indicated a relative abundance of 'same direction copy number changes' in an ICSI setting, with an increase by a factor of five compared to the control population. Nevertheless, due to the limited size of the study group, further research with a larger cohort is necessary to validate these findings [20]. Significantly higher rates of *de novo* chromosomal abnormalities have been observed in ICSI-mediated offspring. These abnormalities include sex chromosome aneuploidies and structural chromosome anomalies, particularly reciprocal translocations [65]. The incidence of *de novo* fetal karyotype abnormalities was determined to be 3.2%, primarily attributed to a substantial number of *de novo* autosomal abnormalities rather than sex chromosome abnormalities. Furthermore, the incidence of *de novo* chromosomal abnormalities, reported at 1.9% in a pooled sample of ICSI offspring assessed both pre- and postnatally, is notably higher compared to the general population and is associated with the paternal sperm count. Higher rates of *de novo* chromosomal abnormalities have been found in the fetuses and children of couples in which men exhibited below-reference sperm concentration and total sperm count. It is important to note that the frequency of abnormal *de novo* prenatal test results did not demonstrate a statistically significant difference when comparing the use of ejaculated versus non-ejaculated sperm or between fresh and frozen-thawed sperm. However, the absence of a control group in the study may have had an impact on the observation of a higher prevalence of karyotype abnormalities. Additionally, it is widely recognized that pregnancies achieved through ART undergo more comprehensive prenatal karyotype screening compared to those conceived naturally. Consequently, to investigate whether the higher incidence of karyotype anomalies following ICSI is a result of closer attention, the authors suggest that the optimal control group should consist of spontaneously conceived pregnancies that adhere to simi-

Table 1. Investigations into ART and *de novo* mutations

Case group (n)*	Control group (n)	Relevant outcomes of the study	Study
12 ICSI	30 SP	In six out of the twelve children who had undergone ICSI, ten apparent *de novo* "genomic copy number changes in one direction" were identified, characterized by simultaneous gains or losses in the copy number relative to both biological parents, with a particular emphasis on losses.	[20]
25 trios ARTP	693 trios SP	The application of ART had a moderate significance ($P = 3.86 \times 10^{-3}$), with an average increase of 4.25 DNMs compared to natural conception after controlling for other variables.	[56]
49 quarters[2] IVF	62 quarters[2] SP	*De novo* numerical aberrations or large structural DNA imbalances occur at comparable frequencies in IVF and those conceived naturally.	[57]
2505 ICSI: 1114 fetuses prenatal tests, 1391 fetuses postnatal tests	–	The incidence of *de novo* chromosomal abnormalities in the pre-and postnatal karyotypes of ICSI offspring was found to be higher compared to the general population, and this increase was associated with the sperm parameters of the fathers.	[60]
116 trios, 86 quartets (twin offspring) ARTP	205 trios SP	Children conceived via ART exhibited an average of 4.59 more gDNMs compared to those conceived spontaneously. This difference includes 3.32 paternal and 1.26 maternal DNMs, after adjusting for factors such as parental age at conception, cigarette smoking, alcohol consumption, and exercise load.	[6]
49 trios and 2 quartets ARTP (17 IVF, 18 ICSI-TESE+ICSI-TESE)	18 trios SP	No significant differences were observed in the number of DNMs per child across various methods of conception, regardless of the parental age at the time of conception.	[58]
1496 ARTP (IVF/ICSI)	1396 SP	The IVF/ICSI group exhibited a slight increase in the overall rate of *de novo* chromosomal abnormalities compared to the NC group; however, this difference did not reach a statistical significance.	[62]
535 ARTP with suspected genetic conditions	1316 SP with suspected genetic conditions	The proportions of *de novo* variants observed in the ART group were comparable to those in the non-ART group.	[59]
270 ARTP, 67 ARTP mother–child pairs, 113 oocytes were donated in natural menstrual cycles	181 SC, 90 SC mother–child pairs, 113 oocytes donated after OS	ART individuals have a higher prevalence of *de novo* non-synonymous mtDNA variants compared to their counterparts conceived spontaneously.	[61]

Trios – mother, father, and child.
Quartets[1] – mother, father, and two siblings.
Quarters[2] – father, mother, placenta, and neonate umbilical cord blood.
ARTP – ART pregnancies (IVF + ICSI).
SP – spontaneous pregnancies (natural cohort).
IVF – in vitro fertilization.
ICSI – intracytoplasmic sperm injection.
TESE-ICSI – testicular sperm extraction associated with intracytoplasmic sperm injection.
*The participants were not diagnosed with any recognized genetic disorders unless so indicated.

lar screening practices, thereby facilitating a more rigorous analysis [60].

Conversely, other research has not identified a correlation between the number of DNMs and ART. Hunag et al. conducted an investigation into the type of molecular defects present in neonates conceived via ART who were admitted to the neonatal intensive care unit with suspected genetic conditions. The genetic profiles of these neonates were compared to those of neonates conceived without ART. The study found no significant differences in the genetic profiles between the two groups, which encompassed the rates of diagnosis and the proportion of *de novo* variants (DNVs), as well as the percentage of SNVs and CNVs between the two cohorts. That study presents several limitations. As noted by the authors, the research focused exclusively on live-born children, thereby excluding stillborn fetuses and pregnancies that had been terminated. This omission may affect the assessment of certain fetal lethal SNVs or CNVs. Additionally, well-documented potential risk factors, such as advanced paternal age, unhealthy parental lifestyles, and infertility, were not examined [59]. The findings align with those by Smits et al., who revealed no significant differences in the number of DNVs between 53 children and their parents [58]. However, the study conducted by Smits and colleagues did not explore the impact of ART on the prevalence of somatic DNMs in the offspring or the occurrence of CNVs. Esteki et al. investigated the prevalence and extent of *de novo* large CNVs greater than 100 kb in live-born neonates conceived through IVF compared to those conceived naturally. The researchers profiled the genomic landscape of fetal and placental tissues postpartum to assess the prevalence and persistence of large genetic aberrations that may be attributed to the chromosomal instability (CIN) associated with IVF. The findings revealed that approximately 10.8% of pregnancies exhibited sporadic *de novo* genomic aberrations distributed throughout the genome, indicating that these events are random and infrequent occurrences of embryonic CIN, with no discernible functional implications for placental biology or fetal health [57]. It is important to highlight that ART is used to mitigate the risk of pathogenic *de novo* mutations, which are more prevalent due to increased mosaicism in parents. This condition elevates the likelihood of recurrence, thereby increasing the risk that a subsequent child may be affected by the same DNM as the preceding one [66].

While our primary focus has been on ART procedures on humans, it is important to acknowledge that research conducted on animal models has also yielded conflicting results. For example, studies involving transgenic mice have demonstrated no significant difference in the frequency or spectrum of DNMs between naturally conceived fetuses and those produced through various ART techniques [67]. Nevertheless, there exist additional data points to consider. The genealogy of domestic cattle is meticulously documented across numerous generations and is characterized by extensive paternal and maternal half-sibling pedigrees. This comprehensive record-keeping presents unique opportunities to construct large pedigrees that are particularly well-suited to the identification of DNMs. Furthermore, the reproduction of domestic cattle extensively employs ART, which includes methods such as artificial insemination (utilizing frozen semen for female insemination), IVF, and ICSI. These techniques facilitate the examination of the impact of ART on DNMs.

When examining species such as cattle, which are particularly suited for investigating this phenomenon, it becomes evident that the correlation between DNMs and the use of ART is significant. Notably, IVF has been shown to increase the rate of *de novo* structural variation (dnSV) by approximately fivefold, with the associated mutations occurring during the very early stages of embryonic development, specifically during the one- and two-cell stages [68]. While ART procedures contribute to the appearance of DNMs, their impact has been shown to be less significant compared to parental aging and other environmental factors (summarized in *Fig. 1*). In summary, while the research conducted in this field has several limitations, the findings related to cattle suggest that further investigations are necessary.

AGE AS A LIMITING FACTOR OF ART

One of the most contentious issues within this context is the consideration of parental age as a basis for restricting access to ART. For instance, in their publication, Bewley et al. emphasized the importance of providing women with comprehensive information regarding reproduction and fertility, as well as the limitations of ART as a means to achieving pregnancy at an advanced age. While their discourse primarily focuses on the impact of biological factors on fertility and the prospects for a healthy pregnancy, they ultimately convey the notion that there exists a specific chronological timeframe during which childbearing is most advisable. Indeed, their subtitle asserts that the most secure age for childbearing remains 20–35 years, against the potential risks associated with ART for women exceeding this age range [69].

The age restrictions on the application of ART vary among countries. For instance, in the United

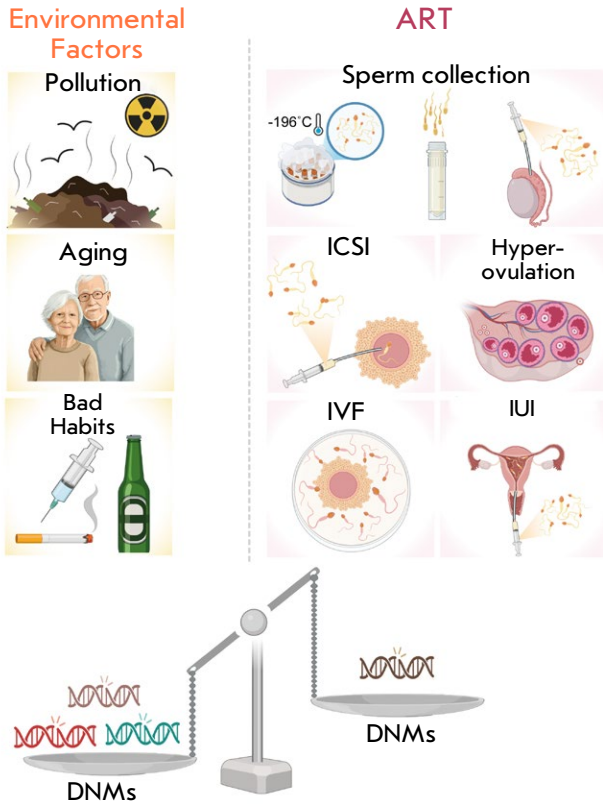


Fig. 1. Factors contributing to *de novo* mutations (DNMs). Two categories of factors that influence the accumulation of DNMs are illustrated: environmental factors (left panel) and assisted reproductive technologies (ART) (right panel)

States, a notable exception to the general absence of age restrictions in statutory law is reflected in the guidelines issued by the Ethics Committee of the American Society for Reproductive Medicine. This committee has explicitly recommended against providing donor oocytes or embryos to women over the age of 55, even in the absence of any underlying medical conditions. The authors justify the establishment of the age limit of 55 by citing the limited data available regarding maternal and fetal safety, as well as concerns related to longevity and the necessity for sufficient psychosocial support for raising a child to adulthood.

The legislation regulating ART throughout Europe reveals that most of the countries have established a chronological age limit for potential mothers, typically ranging from 45 to 50 years. For instance, the Czech Republic has set the age limit for women seeking access to ART at 49 years [70]. In Belgium, prospective mothers must be under 45 years of age to request access to ART and the implantation of an embryo or insemination is not permitted after the age of 48 years [71]. A report from the Swedish National Council on Medical Ethics indicates that there are no mandated upper age limits for IVF in Sweden. However, it is noted that county councils typically impose upper age limits ranging from 37 to 41 years. The report asserts that at least one parent must be of a sufficiently young age to assume responsibility for the child until the child reaches adulthood [72]. This assertion aligns with the Swiss position, which stipulates that both prospective parents should be of an age that enables them to adequately care for their child until the child reaches the age of majority [71]. Estonia, Greece, and the Netherlands set the maternal age limit at 50 years [70]. Notably, a social study investigating public perceptions regarding maternal age revealed that 43 percent of participants believed that women should stop bearing children as they reach the age of 50 [73].

It is noteworthy that, despite the increased risks of health issues in offspring closely associated with advanced paternal age, fathers generally face fewer restrictions regarding the application of ART. In Sweden, prospective fathers are not permitted to be older than 56 years, while in Finland and Portugal, the age limit is set at 60 years [71]. A recent social study performed in the United States revealed that the preferred upper age limit for men was 64 years [74].

Based on an analysis of scientific studies and legislative approaches across various countries, it can be concluded that parental age is a significant factor influencing access to ART, particularly for women. Numerous studies and governmental regulations underscore the risks associated with advanced maternal age, which may lead to potential complications for both the mother and the child. Despite the increasing restrictions imposed, it is crucial to inform prospective parents that successful conception at an advanced age still does not preclude serious implications for the child's health such as an elevated likelihood of DNMs. Consequently, the focus should extend beyond merely regulating access to ART; it should also encompass the provision of comprehensive information regarding potential risks and their implications for the family.

CONCLUSIONS AND FUTURE PERSPECTIVES

The impact of ART as relates to DNMs in humans remains a contentious issue; nevertheless, data show that aging, and paternal aging in particular, exerts

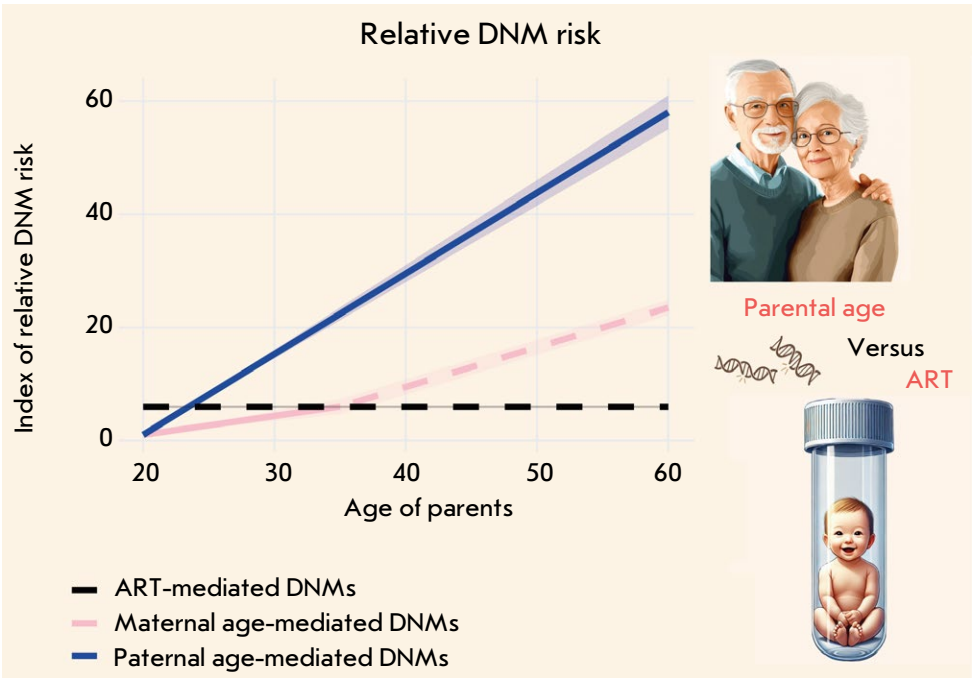


Fig. 2. The impact of parental age and ART on the relative risk of *de novo* mutations (DNMs). The X axis represents parental age, and the Y axis shows the index of relative risk for DNMs, normalized to the minimal risk level observed at age 20. An increase in the index indicates an elevated likelihood of *de novo* mutations. The shaded areas surrounding the lines represent the range of potential risks, which is influenced by the variability of the coefficients used in calculations. The dashed segment of the maternal DNMs line beyond age 35 indicates data revealing a pronounced acceleration in risk, although this increase remains unverified [56, 68].

The following formulas were used for calculating the index:

$\Delta_{paternal}$: The paternal age contribution to DNMs risk:

$\Delta_{paternal} = (Age - 20) \times k_{paternal}$, where $k_{paternal} = 1.35–1.5$ [41–47].

$\Delta_{maternal}$: The maternal age contribution to DNMs risk:

For ages up to 35:

$\Delta_{maternal} = (Age - 20) \times k_{maternal\_low}$, where $k_{maternal\_low} = 0.24–0.42$ [41–47].

For ages above 35:

$\Delta_{maternal} = \Delta_{maternal}(35) + (Age - 35) \times k_{maternal\_high}$, where $\Delta_{maternal}(35)$ is the risk value at age 35, calculated using the first formula, and $k_{maternal\_high} = 0.61$ [42].

The risk associated with ART was calculated as a fixed additional contribution that is added to the overall parental risk, regardless of age, and amounts to ~ 4.25 [56]

the most consequential influence on the number of DNMs, alongside reproductive disorders linked to environmental factors (*Fig. 2*). However, the limitations that come with the small sample sizes of most studies and the predominant use of ART by couples experiencing infertility may obscure the direct impact of the technique as relates to DNMs. Moreover, many existing studies fail to account for stillbirths and pregnancy terminations, which complicates the evaluation of lethal DNMs. To address these limitations, research on model organisms such as cattle, as well as investigations into the mutations in ART-conceived children using various culture media and protocols, could provide valuable insights. This review primarily examined the presence of an association between DNMs and ART, which complicates the evaluation of lethal DNMs epigenetic changes, thus requiring further detailed exploration. Through the examination of these studies, we aimed to gain insights regarding implications for future generations and underscore the growing need to investigate the long-term consequences of ART. ●

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

This work was supported by the grant of the state program of the Sirius Federal Territory "Scientific and Technological Development of the Sirius Federal Territory" (Agreement No. 18-03 dated September 10, 2024, J. Vasilevska).

The figures in this article were created using BioRender.com. Additionally, certain elements of the figures were generated with the support of artificial intelligence tools, including DALL·E and MidJourney. We acknowledge the contributions of these platforms to the visual representation of this work.

REFERENCES

1. ESHRE, 2023. ART fact sheet, data 2019 [Press release]. https://www.eshre.eu/Press-Room/Resources.
2. Wyns C., De Geyter C., Calhaz-Jorge C., Kupka M.S., Motrenko T., Smeenk J., Bergh C., Tandler-Schneider A., Rugescu I.A., et al. // Hum. Reprod. Open. 2022. V. 2022. № 3. P. hoac022. doi: 10.1093/hropen/hoac022.
3. Zegers-Hochschild F., Adamson G.D., Dyer S., Racowsky C., de Mouzon J., Sokol R., Rienzi L., Sunde A., Schmidt L., Cooke I.D., et al. // Fertil Steril. 2017. V. 108. № 3. P. 393–406. doi: 10.1016/j.fertnstert.2017.06.005.
4. Ahmadi H., Aghebati-Maleki L., Rashidiani S., Csabai T., Nnaemeka O.B., Szekeres-Bartho J. // Internat. J. Mol. Sci. 2023. V. 24. № 17. P. 13564. doi: 10.3390/ijms241713564.
5. Goldmann J.M., Veltman J.A., Gilissen C. // Trends Genet. 2019. V. 35. № 11. P. 828–839. doi: 10.1016/j.tig.2019.08.005.
6. Wang C., Lv H., Ling X., Li H., Diao F., Dai J., Du J., Chen T., Xi Q., Zhao Y., et al. // Cell Res. 2021. V. 31. № 8. P. 919–928. doi: 10.1038/s41422-021-00521-w.
7. Heber M.F., Ptak G.E. // Biol. Reprod. 2021. V. 104. № 4. P. 734–744. doi: 10.1093/biolre/ioaa224.
8. Jain M., Singh M. // StatPearls. Treasure Island (FL): StatPearls Publ. 2023. Bookshelf ID: NBK576409.
9. Chamayou S., Giacone F., Cannarella R., Guglielmino A. // J. Clin. Med. 2023. V. 12. № 2. P. 671. doi: 10.3390/jcm12020671.
10. Berntsen S., Laivuori H., la Cour Freiesleben N., Loft A., Söderström-Anttila V., B Oldereid N., et al. // Hum. Reprod. Update. 2021. V. 27. P. 801–847. doi: 10.1093/humupd/dmab005.
11. Zhu J.L., Basso O., Obel C., Bille C., Olsen J. // BMJ. 2006. V. 333. № 7570. P. 679. doi: 10.1136/bmj.38919.495718.AE.
12. Hansen M., Kurinczuk J.J., Bower C., Webb S. // N. Engl. J. Med. 2002. V. 346. № 10. P. 725–730. doi: 10.1056/NEJMoa010035.
13. Rumbold A.R., Moore V.M., Whitrow M.J., Oswald T.K., Moran L.J., Fernandez R.C., Barnhart K.T., Davies M.J. // Hum. Reprod. 2017. V. 32. № 7. P. 1489–1507. doi: 10.1093/humrep/dex085.
14. Wijs L.A., Fusco M.R., Doherty D.A., Keelan J.A., Hart R.J. // Hum. Reprod Update. 2021. V. 28. № 1. P. 132–148. doi: 10.1093/humupd/dmab031.
15. Zhang Z., Liu X., Wei C., Luo J., Shi Y., Lin T., He D., Wei G. // J. Ped. Urol. 2021. V. 17. № 1. P. 9–20. doi: 10.1016/j.jpurol.2020.11.005.
16. Hart R.J., Wijs L.A. // Front. Reprod. Hlth. 2022. V. 4. P. 1045762. doi: 10.3389/frph.2022.1045762.
17. Giorgione V, Parazzini F, Fesslova V, Cipriani S, Candiani M, Inversetti A, Sigismondi C, Tiberio F, Cavoretto P. // Ultrasound Obstet Gynecol. 2018. V. 51. № 1. P. 33-42. doi: 10.1002/uog.18932.
18. La Rovere M., Franzago M., Stuppia L. // Int. J. Mol. Sci. 2019. V. 20. № 17. P. 4169. doi: 10.3390/ijms20174169.
19. Briana D.D., Malamitsi-Puchner A. // J. Matern. Fetal Neonatal Med. 2019. V. 32. № 13. P. 2256–2261. doi: 10.1080/14767058.2018.1425834.
20. Woldringh G.H., Janssen I.M., Hehir-Kwa J.Y., van den Elzen C., Kremer J.A.M., de Boer P., Schoenmakers E.F.P.M. // Hum. Reprod. 2009. V. 24. № 1. P. 233–240. doi: 10.1093/humrep/den323.
21. Heber M.F., Ptak G.E. // Biol. Reprod. 2021. V. 104. № 4. P. 734–744. doi: 10.1093/biolre/ioaa224.
22. Catford S.R., Halliday J., Lewis S., O'Bryan M.K., Handelsman D.J., Hart R.J., McBain J., Rombauts L., Amor D.J., Saffery R., et al. // Fertility Sterility. 2022. V. 117. № 4. P. 727–737. doi: 10.1016/j.fertnstert.2021.12.026.
23. Cannarella R., Crafa A., Mongioì L.M., Leggio L., Iraci N., La Vignera S., Condorelli R.A., Calogero A.E. // J. Clin. Med. 2022. V. 11. № 17. P. 5056. doi: 10.3390/jcm11175056.
24. Koeck R.M., Busato F., Tost J., Consten D., van Echten-Arends J., Mastenbroek S., Wurth Y., Remy S., Langie S., Nawrot T.S., et al. // NPJ Genom. Med. 2022. V. 7. № 1. P. 1–11. doi: 10.1038/s41525-022-00310-3.
25. Novakovic B., Lewis S., Halliday J., Kennedy J., Burgner D.P., Czajko A., Kim B., Sexton-Oates A., Juonala M., Hammarberg K., et al. // Nat. Commun. 2019. V. 10. № 1. P. 3922. doi: 10.1038/s41467-019-11929-9.
26. Sciorio R., Manna C., Fauque P., Rinaudo P. // J. Clin. Med. 2023. V. 12. № 13. P. 4444. doi: 10.3390/jcm11082135.
27. Lopes-Marques M., Mort M., Carneiro J., Azevedo A., Amaro A.P., Cooper D.N., Azevedo L. // Hum. Genom. 2024. V. 18. № 1. P. 20. doi: 10.1186/s40246-024-00587-8.
28. Mohiuddin M., Kooy R.F., Pearson C.E. // Front Genet. 2022. V. 13. P. 983668. doi: 10.3389/fgene.2022.983668.
29. Awadalla P., Gauthier J., Myers R.A., Casals F., Hamdan F.F., Griffing A.R., Côté M., Henrion E., Spiegelman D., Tarabeux J., et al. // Am. J. Hum. Genet. 2010. V. 87. № 3. P. 316–324. doi: 10.1016/j.ajhg.2010.07.019.
30. Satterstrom F.K., Kosmicki J.A., Wang J., Breen M.S., De Rubeis S., An J.-Y., Peng M., Collins R., Grove J., Klei L., et al. // Cell. 2020. V. 180. № 3. P. 568–584. doi: 10.1016/j.cell.2019.12.036.
31. An J.-Y., Lin K., Zhu L., Werling D.M., Dong S., Brand H., Wang H.Z., Zhao X., Schwartz G.B., Collins R.L., et

al. // Science. 2018. V. 362. № 6420. P. 6576. doi: 10.1126/science.aat6576.
32. Jin S.C., Homsy J., Zaidi S., Lu Q., Morton S., DePalma S.R., Zeng X., Qi H., Chang W., Sierant M.C., et al. // Nat. Genet. 2017. V. 49. № 11. P. 1593–1601. doi: 10.1038/ng.3970.
33. Jin Z.-B., Wu J., Huang X.-F., Feng C.-Y., Cai X.-B., Mao J.-Y., Xiang L., Wu K.-C., Xiao X., Kloss B.A., et al. // Proc. Natl. Acad. Sci. USA. 2017. V. 114. № 16. P. 4219–4224. doi: 10.1073/pnas.1615970114.
34. Veltman J.A., Brunner H.G. // Nat. Rev. Genet. 2012. V. 13. № 8. P. 565–575. doi: 10.1038/nrg3241.
35. Li Q., Zhao L., Zeng Y., Kuang Y., Guan Y., Chen B., Xu S., Tang B., Wu L., Mao X., et al. // Genome Biol. 2023. V. 24. № 1. P. 68. doi: 10.1186/s13059-023-02894-0.
36. Oud M.S., Smits R.M., Smith H.E., Mastrorosa F.K., Holt G.S., Houston B.J., de Vries P.F., Alobaidi B.K.S., Batty L.E., Ismail H., et al. // Nat. Commun. 2022. V. 13. № 1. P. 154. doi: 10.1038/s41467-021-27132-8.
37. Acuna-Hidalgo R., Veltman J.A., Hoischen A. // Genome Biol. 2016. V. 17. № 1. P. 241. doi: 10.1186/s13059-016-1110-1.
38. Zhao G., Li K., Li B., Wang Z., Fang Z., Wang X., Zhang Y., Luo T., Zhou Q., Wang L., et al. // Nucl. Acids Res. 2020. V. 48. № D1. P. D913–D926. doi: 10.1093/nar/gkz923.
39. Turner T.N., Yi Q., Krumm N., Huddleston J., Hoekzema K.F., Stessman H.A., Doebley A.-L., Bernier R.A., Nickerson D.A., Eichler E.E. // Nucl. Acids Res. 2017. V. 45. № D1. P. D804–D811. doi: 10.1093/nar/gkw865.
40. Campbell C.D., Eichler E.E. // Trends Genet. 2013. V. 29. № 10. P. 575–584. doi: 10.1016/j.tig.2013.04.005.
41. Sasani T.A., Pedersen B.S., Gao Z., Baird L., Przeworski M., Jorde L.B., Quinlan A.R. // eLife. 2019. V. 8. P. e46922. doi: 10.7554/eLife.46922.
42. Wong W.S.W., Solomon B.D., Bodian D.L., Kothiyal P., Eley G., Huddleston K.C., Baker R., Thach D.C., Iyer R.K., Vockley J.G., et al. // Nat. Commun. 2016. V. 7. P. 10486. doi: 10.1038/ncomms10486.
43. Kong A., Frigge M.L., Masson G., Besenbacher S., Sulem P., Magnusson G., Gudjonsson S.A., Sigurdsson A., Jonasdottir A., Jonasdottir A., et al. // Nature. 2012. V. 488. № 7412. P. 471. doi: 10.1038/nature11396.
44. Wood K.A., Goriely A. // Fertility Sterility. 2022. V. 118. № 6. P. 1001. doi: 10.1016/j.fertnstert.2022.10.017.
45. Goldmann J.M., Veltman J.A., Gilissen C. // Trends Genet. 2019. V. 35. № 11. P. 828–839. doi: 10.1038/ng.3597.
46. Jónsson H., Sulem P., Kehr B., Kristmundsdottir S., Zink F., Hjartarson E., Hardarson M.T., Hjorleifsson K.E., Eggertsson H.P., Gudjonsson S.A., et al. // Nature. 2017. V. 549. № 7673. P. 519–522. doi: 10.1038/nature24018.
47. Kessler M.D., Loesch D.P., Perry J.A., Heard-Costa N.L., Taliun D., Cade B.E., Wang H., Daya M., Ziniti J., Datta S., et al. // Proc. Natl. Acad. Sci. USA. 2020. V. 117. № 5. P. 2560–2569. doi: 10.1073/pnas.1902766117.
48. Xia B., Yan Y., Baron M., Wagner F., Barkley D., Chiodin M., Kim S.Y., Keefe D.L., Alukal J.P., Boeke J.D., et al. // Cell. 2020. V. 180. № 2. P. 248–262.e21. doi: 10.1016/j.cell.2019.12.015.
49. Choufani S., Turinsky A.L., Melamed N., Greenblatt E., Brudno M., Bérard A., Fraser W.D., Weksberg R., Trasler J., Monnier P., et al. // Hum. Mol. Genet. 2019. V. 28. № 3. P. 372–385. doi: 10.1093/hmg/ddy321.
50. Aitken R.J. // Hum. Reprod. 2023. V. 38. № 10. P. 1861–1871. doi: 10.1093/humrep/dead157.
51. Gao Z., Moorjani P., Sasani T.A., Pedersen B.S., Quinlan A.R., Jorde L.B., Amster G., Przeworski M. // Proc. Natl. Acad. Sci. USA. 2019. V. 116. № 19. P. 9491–9500. doi: 10.1073/pnas.1901259116.
52. Aitken R.J. // Reprod. BioMed. Online. 2022. V. 45. № 1. P. 109–124. doi: 10.1016/j.rbmo.2022.03.012.
53. Crow J.F. // Nat. Rev. Genet. 2000. V. 1. № 1. P. 40–47. doi: 10.1038/35049558.
54. Jónsson H., Sulem P., Arnadottir G.A., Pálsson G., Eggertsson H.P., Kristmundsdottir S., Zink F., Kehr B., Hjorleifsson K.E., Jensson B.Ö., et al. // Nat. Genet. 2018. V. 50. № 12. P. 1674–1680. doi: 10.1038/s41588-018-0259-9.
55. Belyeu J.R., Brand H., Wang H., Zhao X., Pedersen B.S., Feusier J., Gupta M., Nicholas T.J., Brown J., Baird L., et al. // Am. J. Hum. Genet. 2021. V. 108. № 4. P. 597–607. doi: 10.1016/j.ajhg.2021.02.012.
56. Wong W.P., Hassed C., Chambers R., Coles J. // Front Aging Neurosci. 2016. V. 8. P. 156. doi: 10.3389/fnagi.2016.00156.
57. Zamani Esteki M., Viltrop T., Tšuiko O., Tiirats A., Koel M., Nõukas M., Žilina O., Teearu K., Marjonen H., Kahila H., et al. // Nat. Med. 2019. V. 25. № 11. P. 1699–1705. doi: 10.1038/s41591-019-0620-2.
58. Smits R.M., Xavier M.J., Oud M.S., Astuti G.D.N., Meijerink A.M., de Vries P.F., Holt G.S., Alobaidi B.K.S., Batty L.E., Khazeeva G., et al. // Hum. Reprod. 2022. V. 37. № 6. P. 1360–1369. doi: 10.1093/humrep/deac068.
59. Huang Z., Xiao F., Xiao H., Lu Y., Yang L., Zhuang D., Chen L., Wei Q., Jiang Y., Li G., et al. // JAMA Netw Open. 2023. V. 6. № 4. P. e236537. doi: 10.1001/jamanetworkopen.2023.6537.
60. Belva F., Bonduelle M., Buysse A., van den Bogaert A., Hes F., Roelants M., Verheyen G., Tournaye H., Keymolen K. // Hum. Reprod. 2020. V. 35. № 9. P. 2149–2162. doi: 10.1093/humrep/deaa162.
61. Mertens J., Belva F., van Montfoort A.P.A., Regin M., Zambelli F., Seneca S., Couvreu de Deckersberg E., Bonduelle M., Tournaye H., Stouffs K., et al. // Nat. Commun. 2024. V. 15. № 1. P. 1232. doi: 10.1038/s41467-024-45446-1.
62. Yuan S., Guo L., Cheng D., Li X., Hu H., Hu L., Lu G., Lin G., Gong F., Tan Y.-Q. // J. Assist. Reprod. Genet. 2022. V. 39. № 7. P. 1683–1689. doi: 10.1007/s10815-022-02500-5.
63. Sciorio R., Esteves S.C. // J. Clin. Med. 2022. V. 11. № 8. P. 2135. doi: 10.3390/jcm11082135.
64. Oud M.S., Smits R.M., Smith H.E., Mastrorosa F.K., Holt G.S., Houston B.J., de Vries P.F., Alobaidi B.K.S., Batty L.E., Ismail H., et al. // Nat. Commun. 2022. V. 13. № 1. P. 154. doi: 10.1038/s41467-021-27132-8.
65. Bonduelle M., van Assche E., Joris H., Keymolen K., Devroey P., Van Steirteghem A., Liebaers I. // Hum. Reprod. 2002. V. 17. № 10. P. 2600–2614. doi: 10.1093/humrep/17.10.2600.
66. Xu N., Shi W., Cao X., Zhou X., Jin L., Huang H.-F., Chen S., Xu C. // J. Med. Genet. 2023. V. 60. № 9. P. 910–917. doi: 10.1136/jmg-2022-108920.
67. Caperton L., Murphey P., Yamazaki Y., McMahan C.A., Walter C.A., Yanagimachi R., McCarrey J.R. // Proc. Natl. Acad. Sci. USA. 2007. V. 104. № 12. P. 5085–5090. doi: 10.1073/pnas.0611642104.
68. Lee Y.-L., Bouwman A.C., Harland C., Bosse M., Costa Monteiro Moreira G., Veerkamp R.F., Mullaart E., Cambisano N., Groenen M.A.M., Karim L., et al. // Genome Res. 2023. V. 33. № 9. P. 1455–1464. doi: 10.1101/gr.277884.123.
69. Martin L.J. // Women's Stud. Internat. Forum. 2017. V. 62. P. 91–98. doi: 10.1016/j.wsif.2017.04.004.
70. Bewley S., Davies M., Braude P. // BMJ. 2005. V. 331. № 7517. P. 588–589. doi: 10.1136/bmj.331.7517.588.

71. Kocourková J., Konečná H., Burcin B., Kučera T. // Reprod. Biomed. Online. 2015. V. 30. № 5. P. 482–492. doi: 10.1016/j.rbmo.2015.01.017.
72. Martani A., De Clercq E., De Geyter C., Pennings G., Wangmo T., Elger B.S. // J. Law Biosci. 2022. V. 9. № 2. P. lsac036. doi: 10.1093/jlb/lsac036.
73. Londra L., Wallach E., Zhao Y. // Semin. Fetal Neonat. Med. 2014. V. 19. № 5. P. 264–271. doi: 10.1016/j.siny.2014.07.003.
74. Bowman M.C., Saunders D.M. // Hum. Reprod. 1994. V. 9. № 1. P. 167–171. doi: 0.1093/oxfordjournals.humrep.a138309.

Generation of TIL-based Cellular Products for Cancer Immunotherapy: Current Insights and the Challenges

D. V. Kuznetsova*, T. V. Petrova
Lopukhin Federal Research and Clinical Center of Physical–Chemical Medicine, Moscow, 119435 Russia
*E-mail: dashakuz23@gmail.com
Received November 26, 2024; in final form, February 11, 2025
DOI: 10.32607/actanaturae.27559

ABSTRACT **Tumor-infiltrating T lymphocytes (TILs) are a population of T cells present in tumor tissue and enriched in tumor antigen-specific clones. TILs participate in the adaptive antitumor immune response, which makes them a promising candidate for cancer immunotherapy. The concept framing this type of therapy involves the extraction of T cells from a patient's tumor, followed by their *in vitro* expansion and reinfusion into the same patient in large quantities. This approach enhances the antitumor immune response and allows one to affect cancer cells resistant to other types of treatment. In 2024, the first TIL-based drug was approved for melanoma treatment. The possibility of using TILs for treating other solid tumors is currently being considered, and novel methods aiming to increase the efficiency of generating TIL cultures from tumor tissues *in vitro* are being developed. However, despite the significant progress achieved in this area, there remain unresolved issues and problems, including the lack of standardized protocols for obtaining, expanding, and cryopreserving TILs, the complexity related to their isolation and the duration of that, as well as insufficient efficiency. Our review focuses on the concept of immunotherapy using TILs, the main stages involved in generating a TIL-based cellular product, associated problems, and further steps in the production of TIL cultures that aim to improve efficiency as relates to production and ensure a wider application of the therapy.**
KEYWORDS **tumor-infiltrating T lymphocytes, immunotherapy, T-cell therapy, TIL.**
ABBREVIATIONS **TIL – tumor-infiltrating T lymphocytes; CTLA-4 – cytotoxic T lymphocyte-associated protein 4; PD-1 – programmed cell death protein 1; PD-L1 – programmed death-ligand 1; CAR – chimeric antigen receptor; CAR-T – chimeric antigen receptor to T cells; CAR-NK – chimeric antigen receptor-engineered natural killer; TCR – T-cell receptor; FDA – Food and Drug Administration, pre-REP – pre-rapid expansion protocol; REP – rapid expansion protocol; PGE2 – prostaglandin E2; IL-2 – interleukin-2; IL-2Rα/β – interleukin-2 receptor alpha and beta chain; ortho-hIL-2 –orthogonal human genetically engineered interleukin-2; ortho-hIL-2Rβ – orthogonal human genetically engineered interleukin-2 receptor; IL-7 – interleukin-7; IL-12 – interleukin-12; IL-12Rb1 – interleukin-12 receptor beta 1 subunit; IL-15 – interleukin-15; mbIL15 – membrane-bound interleukin-15; IL-21 – interleukin-21; 4-1BB (CD137 or TNFRSF9) – member of the tumor necrosis factor receptor family; IFN-γ – interferon gamma; PBMC – peripheral blood mononuclear cell; DMSO – dimethyl sulfoxide; Tregs – regulatory T cells; NK cells – natural killer cells; MHC – major histocompatibility complex; pMHC – major histocompatibility complex peptide.**

INTRODUCTION

Cancer immunotherapy is among the most innovation-prolific and promising areas of modern oncology. As new data on the interplay between the immune system and tumors become available, various forms of immunotherapy (therapy using immune checkpoint inhibitors, including antibodies specific to molecules such as CTLA-4, PD-1, PD-L1, etc.; CAR T cell therapy, CAR NK cell therapy; dendritic cell therapy; *in vitro* generation followed by reinfusion of autologous tumor-infiltrating T lymphocytes (TILs) back into the patient's body; and vaccination with chemically

synthesized neoantigen peptides) has started to be viewed as a promising novel approach to the treatment of different types of malignant tumors, since it allows practicians to personalize treatment and improve its efficacy even in patients with uncontrolled and metastatic cancer.

Currently, cancer immunotherapy is fertile ground for the research and development of novel drugs.

Tumor-infiltrating T lymphocytes are a population of T cells within tumor tissue that are enriched in tumor-specific clones. However, the immunosuppressive factors inherent in the tumor microenvironment actively suppress the antitumor immune response and weaken the ability of TILs to destroy tumor cells. The concept of TIL therapy is based on the idea that the antitumor immune response can be restored by isolating TILs from a tumor fragment, culturing them *ex vivo* to increase their quantity (to at least 10^9 cells), and finally reinfusing them back into the patient. Unlike other cell-based immunotherapy methods, TILs are obtained directly from the patient, without any genetic modification [1].

The research in the 1950s that aimed to explore how possible it was to employ T cells to suppress tumor cell growth was inspired by studies that had demonstrated that rejection of solid organ transplants was mediated by cellular immunity [2]. Animal experiments showed that when transferred to syngeneic recipients, T cells from immunized donors could mediate tumor regression, and that IL-2 could be used to increase their number [3]. Later, it was revealed using a mouse model that simultaneous administration of IL-2 and T cells *in vivo* enhances the antitumor efficacy of T cells. However, the requirement that an immunized syngeneic donor be the source of the tumor-specific T cells remained a hurdle in attempts to use this approach in humans lacking such a source of TILs.

This hurdle was overcome in 1986, when Rosenberg and colleagues from the Surgery Branch of the National Cancer Institute (USA) became the first to demonstrate, in a mouse model, that a combination of autologous TILs and cyclophosphamide could induce a regression of metastases [4]. Next came a landmark publication in 1988 that became the first study to show that infusion of TILs into patients with metastatic melanoma could lead to tumor regression [5]. As of October 2024, a total of 266 clinical trials related to TIL therapy had been registered on ClinicalTrials.gov. Of those, 26 trials have the "active" status; 103 trials are "recruiting participants," and 82 trials have been completed [6]. Over the past five years, 15–30 new clinical trials to assess TIL therapy against various solid tumors have been registered annually, melanoma being predominant (40% of all clinical trials) [7].

Table 1 presents a selective list of phases I and II clinical trials that embrace nearly all solid tumor types.

On February 16, 2024, the FDA approved lifileucel (Amtagvi), the result of 30 years of research, as the first TIL-based therapeutic. The drug was approved for adult patients with unresectable or metastatic melanoma who had previously received standard treatment. Lifileucel is produced through *ex vivo* cultivation of tumor-infiltrating T lymphocytes derived from surgically resected autologous tumor fragments [8].

Metastatic melanoma is considered a highly immunogenic malignant tumor. The objective response rate to TIL therapy ranges from 36 to 56%; progression-free survival is 3.7–7.5 months; the overall survival time ranges from 15.9 to 21.8 months [9]. Less immunogenic (also known as "cold") tumors respond worse to TIL therapy, which poses a problem on one hand, while, on the other hand, it opens up new avenues towards developing new strategies to optimize TIL-based treatments.

EX VIVO PRODUCTION AND EXPANSION OF TUMOR-INFILTRATING T LYMPHOCYTES

Ex vivo expansion of tumor-infiltrating T lymphocytes can be divided into two stages: the production of TIL cultures from tumor tissue (the pre-REP stage) and large-scale expansion of T cells (the REP stage) (*Fig. 1*).

Freshly resected tumor tissue obtained during a surgical resection is promptly transported to the laboratory within several hours after surgery in a sterile container with the transport medium (a growth medium supplemented with an antibiotic). The biological material is immediately cut into small fragments sized approximately 1.5 × 1.5 mm^2 and placed into a growth medium supplemented with interleukin-2 (IL-2) at concentrations ranging from 500 to 6,000 IU/mL. An alternative method for TIL culture generation involves enzymatic digestion of tumor fragments in an enzyme cocktail containing collagenase and DNase at 37°C for 30–60 min. The resulting cell suspension is subsequently transferred to a growth medium supplemented with IL-2 (500–6,000 IU/mL) [10–13]. To further activate TIL cultures, IL-2 is used in combination with anti-CD3/CD28 antibodies in some protocols [14–18]. Since clinical research includes studying the feasibility of producing TILs from tumors of different localizations, including skin and gastrointestinal tumors, one should bear in mind that bacterial contamination of tumor fragments is possible. Therefore,

Table 1. Selected clinical trials of TIL therapy registered on Clinicaltrials.gov as of October 2024

Nosological entity	NCT identifier number	Phase of clinical trial	Number of patients	Administered dose
Stage IIIb, IIIc or IV melanoma	NCT03374839	I/II	11	Cohort 1: 5×10^8 TILs (three patients)/ Cohort 2: 1–20 $\times 10^9$ TILs on weeks 14 and 18
Stage IV melanoma	NCT03475134	I	10	N/A
Measurable metastatic melanoma	NCT03166397	II	30	N/A
Unresectable stage III/IV melanoma or platinum-resistant ovarian cancer	NCT03158935	Ib	24	1×10^{10}–1.6×10^{11} TILs
Unresectable stage III/IV cutaneous or mucosal melanoma	NCT02652455	Pilot	12	N/A, cell growth after 4–8 weeks when using CD137-activating antibody
Measurable metastatic melanoma	NCT02621021	II	170	N/A, young TILs
Unresectable metastatic melanoma	NCT02360579	II	60	N/A
Metastatic melanoma or stage III in-transit, subcutaneous, or regional nodal disease	NCT01740557	Pilot	15	Up to 1.5×10^{11} TILs
Unresectable stage III/IV melanoma	NCT02354690	I/II	12	1×10^9–2×10^{11} TILs
Unresectable stage III/IV melanoma	NCT02278887	III	168	N/A
Metastatic melanoma or stage III in-transit, subcutaneous, or regional nodal disease	NCT01955460	Pilot	15	Up to 1.5×10^{11} TILs
Metastatic melanoma	NCT01993719	II	64	N/A
Unresectable stage III or IV melanoma	NCT01946373	I	10	Up to 5×10^{10} TILs
Unresectable stage III/IV melanoma	NCT01883323	II	12	1×10^{10} –1.6×10^{11} TILs
Metastatic melanoma, uveal melanoma or stage III in-transit or regional nodal disease	NCT00338377	II	189	Cohort 1–3: up to 1.5×10^{11} TILs. Cohort 4: 5.0×10^9 TILs on day 1, 10×10^{10} TILs on day 15
Metastatic uveal melanoma	NCT03467516	II	59	1×10^9–2×10^{11} TILs
Metastatic melanoma	NCT01995344	II	90	N/A
Unresectable stage III/IV melanoma	NCT02379195	I/II	12	N/A
Stage III/IV melanoma	NCT01807182	II	13	N/A
Unresectable melanoma, stage III/IV	NCT01701674	Pilot	13	N/A
Unresectable stage IV metastatic melanoma or stage III in-transit or regional nodal disease	NCT01659151	II	17	N/A
Metastatic melanoma	NCT01319565	II	102	Cohort 1 + 2: 1×10^9 –2×10^{11} young TILs
Unresectable stage III/IV melanoma	NCT01005745	I/II	19	N/A
Locally advanced, recurrent, or metastatic biliary tract cancer	NCT03801083	II	59	2×10^{11} TILs (at least 1×10^9 cells)
Metastatic uveal melanoma	NCT03467516	II	47	2×10^{11} TILs (at least 1×10^9 cells)
Breast cancer	NCT05142475	I	50	1×10^9–5×10^{10} TILs
Malignant solid tumors	NCT05649618	I	42	2.5×10^9–5×10^{10} TILs
Advanced solid cancers	NCT03935893	II	240	2×10^{11} TILs (at least 1×10^9 cells)
Malignant solid tumors	NCT05902520	I	18	N/A
Urothelial cell carcinoma (UCC) and non-muscle invasive bladder urothelial carcinoma (NMIBC)	NCT05768347	I	12	N/A
Advanced melanoma	NCT05098184	I	50	1×10^9–5×10^{10} TILs
Metastatic III and IV stage melanoma	NCT01883323	II	12	1.0×10^6 cells/mL and expanded for no longer than 28 days prior to cryopreservation
Melanoma	NCT02360579	II	66	26.1×10^9 (range, 3.3–72) TILs
Non-small cell lung cancer	NCT04614103	II	170	1×10^9–150×10^9 TILs
Cervical cancer	NCT03108495	II	27	28×10^9 TILs

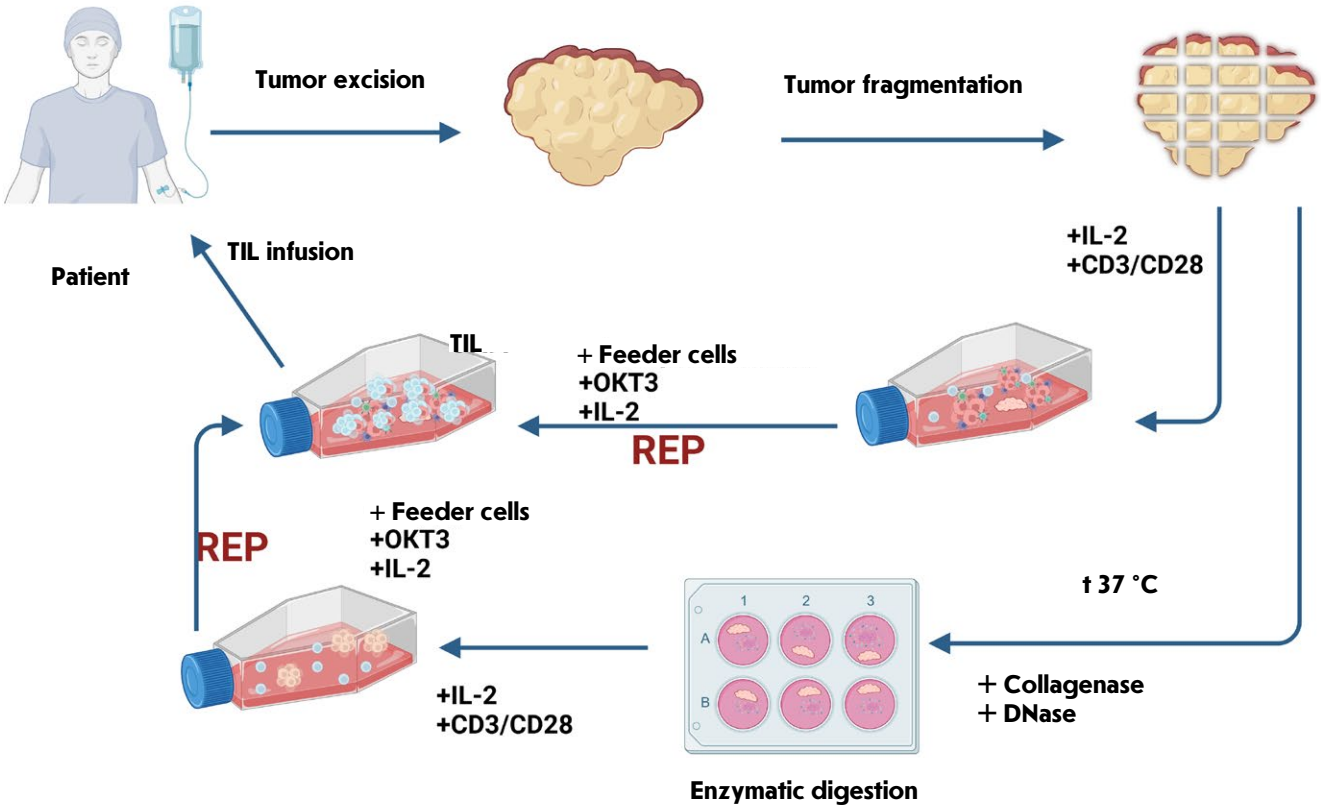


Fig 1. Preparation and infusion of TILs obtained from freshly resected tumor tissue. The two most common options are shown: obtaining TILs from tumor fragments and by enzymatic digestion. Regardless of the type of pre-REP stage of TIL production, at the second stage (REP), feeder cells need to be added to ensure large-scale expansion before infusing the cellular product into the patient

additional washing steps and/or *ex vivo* cultivation in the presence of antibiotics and antifungals are recommended to mitigate this risk. Some protocols also involve pre-incubation of tumor fragments in a medium containing 10% antibiotics at room temperature for 30 min prior to further manipulations, in particular when working with colorectal cancer or melanoma specimens [19]. The initial stage is considered completed once the cell count in the primary TIL culture reaches ~ 10^6 cells per mL of suspension. Next, TILs may either undergo cryopreservation or one can proceed to the second stage: large-scale expansion (REP) aiming to generate a clinically significant number of cells. According to clinical trials and the instruction for use of the approved medicinal product lifileucel, the number of TILs required for infusion ranges from 1×10^9 to 2×10^{11} cells; the total infusion volume being 100–400 mL [12, 20]. Feeder cells (either peripheral blood mononuclear cells from healthy donors (allogeneic) or from the patient (syngeneic)) pre-irradiated with 40 Gy are utilized for large-scale expansion during the second phase. The feeder cells are cocultured with TILs in a growth medium containing IL-2 (500–6,000 IU/mL) until the clinically significant number of TILs is reached [19]. Since effective *ex vivo* TIL expansion largely depends on the number of feeder cells, standard protocols recommend using the 100 : 1 or even 200 : 1 ratio of feeder cells to TILs [21]. It is commonly believed that fewer feeder cells can significantly reduce the yield of TILs, thus underscoring their importance for successful expansion [15]. Because large quantities of feeder cells need to be utilized, clinicians often use donor-derived feeder cells and pool material from multiple donors

[12]. This approach is unique to TIL therapy compared to the more common method used for producing CAR T cells, where feeder cells are not employed. Instead, high doses of IL-2 and anti-CD3/CD28 activating antibodies are simultaneously added directly to the growth medium to stimulate T-cell proliferation [22, 23].

EFFECTIVENESS IN THE GENERATION OF TUMOR-INFILTRATING T LYMPHOCYTES

As mentioned previously, reliable generation of TIL cultures from tumor tissue fragments is the cornerstone of successful TIL therapy. An analysis of the studies conducted by various researchers (*Table 2*) revealed that the likelihood of obtaining viable TIL cultures from patients' tumor fragments is weakly dependent on the type of solid tumor and varies across study sites. TIL cultures were successfully generated in 18–100% of patients across the studies. The findings summarized in *Table 2* infer that this variability is partly attributable to the lack of standardization for the TIL culture generation procedures, as well as to the fact that certain tumor types (e.g., colorectal cancer and melanoma) carry a higher risk of microbiological contamination. Furthermore, factors such as the quantity of initial tumor biomaterial and the degree of immune cell infiltration into it (as observed in uveal melanoma and glioblastoma) play a rather significant role. Unfortunately, the small sample size in most studies weighs negatively on the integrity and validity of the reported data and may lead to both over- and underestimation of the effectiveness of TIL culture generation. We have found just one study that focused on effectiveness in TIL culture generation in a large patient cohort (over 1,000 subjects). It could be inferred from the results of the study that the effectiveness varied by year; over an 11-year period, TIL cultures were produced in an average of < 70% of patients [33].

NOVEL APPROACHES TO THE GENERATION OF TUMOR-INFILTRATING T LYMPHOCYTES

Despite significant progress, effectiveness in TIL culture generation remains well below 100%. Moreover, TIL cultures must be enriched with cytotoxic CD8+ T cells to ensure an optimal antitumor response *in vivo*. Meanwhile, *Table 2* suggests that the proportion of CD8+ T cells greatly varies and is potentially affected by both the initial ratio of T cells within the tumor tissue and the specific culture conditions.

Current research focuses on optimizing protocols for TIL culture generation by supplementing the growth medium with various interleukin cocktails, utilizing genetic engineering at different T-cell production stages, and working with the immunosuppressive tumor microenvironment, which can ruin the full potential of antitumor cellular therapy.

One of the approaches to enhancing effectiveness in *ex vivo* generation of TIL cultures from tumor tissue involves adding immune checkpoint modulators into the growth medium. Several research groups have demonstrated that adding an agonistic anti-4-1BB antibody to melanoma tissue fragments reduces the expansion duration and increases the proportion of CD8+ T cells within the TIL culture compared to a conventional growth medium containing IL-2 only [35, 36]. Similar effects by this antibody have been observed in 16 samples of non-small cell lung cancer. A combination of IL-2 and agonistic anti-CD3 and anti-4-1BB antibodies (urelumab) added to the TIL culture medium reduced the time required to generate TIL cultures and increased the proportion of CD8+ T cells during both the pre-REP and REP stages of TIL culturing in [37]. This approach ensured 100% effectiveness during TIL culture generation for 12 uveal melanoma samples [33]. Since uveal melanoma is characterized by a low immune cell infiltration, TIL culture generation from this tumor type poses a significant challenge. The number of TILs obtained in this study from five fragments less than 3 mm^3 in size was comparable to, or exceeded, that produced from 20 fragments using the conventional method (IL-2 only) [33].

Another approach to the interplay with the tumor microenvironment was proposed by a research team that had demonstrated the effectiveness of inhibiting the prostaglandin E2 (PGE2) signaling pathway to stimulate an antitumor response *in vivo* [38]. Relying on these findings, Morotti et al. discovered that effectiveness in TIL culture generation from melanoma samples (NCT03475134) could be improved by inhibiting the PGE2 signaling pathway. Inhibition of this signaling pathway increased the susceptibility of TILs to IL-2, thus reducing the impact of oxidative stress on T cells and their ferroptosis-mediated death [39].

Addition of various interleukin cocktails to the growth medium is another promising approach to enhancing effectiveness in TIL culture generation, including the production of cultures exhibiting tailored properties (e.g., TIL cultures with a predominant proportion of CD8+ T cells or cultures enriched in memory T cells rather than effector T cells). The application of interleukin cocktails involves a move away from the conventional use of IL-2 alone for T-cell activation and allows one to study how different cytokines (IL-4, IL-7, IL-15, and IL-21) and their combinations affect the end cellular product. Cytokine cocktails had originally been widely used to culture another cellular product: CAR T cells [40–42]. In fur-

ther studies, CAR T cells cultured in media supplemented with IL-7 and IL-15 exhibited higher proliferation rates and enhanced antitumor activity compared to cells cultured just in the presence of IL-2 [43]. Furthermore, it has been established that adding a combination of IL-2, IL-15, and IL-21 increases the CD8+/CD4+ T-cell ratio [44], which is especially important in CAR T therapy.

Because of the successful application of various interleukin combinations in CAR T therapy, similar approaches are now being adopted for the generation of TIL-based products. Studies involving PD-1+CD8+ T cells isolated from the blood of healthy donors and patients with a confirmed diagnosis of cancer demonstrated that a cytokine cocktail containing IL-7 and IL-15 added to the growth medium, along with anti-CD3/CD28 antibodies, significantly enhances T-cell proliferation in the suspension [45]. Treatment with a combination of anti-CD3 antibodies, panobinostat, IL-2, and IL-21 was shown to increase the proportion of CD62L+CD28+CD8+ T cells in TIL cultures compared to TILs cultured in the absence of this cytokine cocktail [46].

The research into the use of interleukins to enhance T-cell expansion continues to advance; modified forms of interleukin are being actively developed. For example, a genetically engineered IL-2 (STK-012) is currently under development; it is the first-in-class partial agonist of the IL-2 receptor alpha and beta chains (IL-2Ra/β) required to selectively activate CD25+ antigen-activated T cells without inducing the nonspecific activation of NK cells or naïve T cells. Preclinical *in vivo* studies in mice using the murine surrogate mSTK-012 revealed a significant reduction in the number of exhausted T cells and increased systemic and intratumoral expansion of the tumor antigen-specific CD25+PD-1+CD8+ T cell population. Additionally, the number of intratumoral regulatory T cells (Tregs) was decreased, indicating that mSTK-012 exhibits better antitumor properties compared to those of IL-2 [47, 48].

Orthogonal cytokine–receptor pairs for human IL-2 that interact exclusively with each other have been developed when studying interleukin modifications. Notably, these pairs do not interact with their native counterparts: cytokine IL-2 and its receptor IL-2. Introduction of orthoIL-2Rβ into the T-cell suspension has enabled selective targeting of orthoIL-2 to genetically modified CD4+ and CD8+ T cells, both *in vitro* and *in vivo*. This approach can reduce adverse events and minimize toxicity compared to that of the canonical form of IL-2 [49].

The next, potential candidate modifier of the antitumor activity of T cells is interleukin-12 (IL-12), a pro-inflammatory cytokine that plays a crucial role in the activation of CD4+ and CD8+ T cells, as well as NK cells. The high toxicity of IL-12 has been limiting its clinical application. Preclinical studies suggest that the toxicity of IL-12 is primarily associated with the activation of NK cells. An attempt was made to address this problem using an IL-12 partial agonist (STK-026), which has reduced affinity for binding to the IL-12 receptor β1 subunit (IL-12Rb1). STK-026 selectively affected activated T cells characterized by upregulated IL-12Rb1 expression, whereas NK cells or resting T cells with moderate IL-12Rb1 expression levels were not significantly affected by STK-026 [50]. The Synthekine company is currently conducting preclinical trials for STK-026, which are expected to demonstrate its capacity to activate tumor-infiltrating CD8+ T cells and myeloid cells, as well as its antitumor efficacy and pharmacodynamic profile.

As mentioned previously, genetic modification of T cells is a possible path in addressing the problem of efficient TIL culture generation and enhancement of their functionality.

Recent studies have shown the great potential that lies in engineering T cells carrying an inducible membrane-bound IL-12. These modified T cells exhibited superior cytotoxic activity *in vitro* and were characterized by a significant level of IFN-γ production [51].

Obsidian Therapeutics, a pharmaceutical company, is currently involved in a multicenter clinical trial to evaluate potential uses for genetically modified TILs OBX-115 expressing membrane-bound IL-15 (mbIL15). This approach allows one to avoid *in vivo* administration of high-dose IL-2, thereby reducing the toxicity and expanding the applicability of TIL therapy to larger patient cohorts [52].

Rejuvenation of tumor-infiltrating T lymphocytes is another interesting strategy for augmenting their antitumor activity using genetic engineering means. This approach allows for the rejuvenation of TILs by restoring their original functionality and potential via partial reprogramming using transient expression of a set of transcription factors. The rejuvenated TILs retain a diverse repertoire of their T-cell receptors (TCRs), thus ensuring broad antigenic specificity. The key positives of TIL rejuvenation consist in a reduction of the epigenetic age of T cells, higher expansion rates, acquisition of a stem cell phenotype, and increased cytokine secretion upon activation by target antigens. Importantly, positive results have been achieved not only for rejuvenated TILs but also for rejuvenated peripheral blood mononuclear cells (PBMCs), TCR and CAR T cells, which indicates that the rejuvenation technology can be widely applied in cancer immunotherapy [51].

Table 2. The features and effectiveness of generating TIL cultures from solid tumors

Nosological entity	State of the initial tumor tissue sample	Number of tumor tissue samples in the study	Features of TIL culture generation	Effectiveness of TIL culture generation, %	Percentage of CD4+, CD8+ T- cell populations of the total cell count in the TIL culture, %	Percentage of contaminated TIL cultures	Reference
Melanoma	Freshly resected material	90 tumors, 710 individual cultures	1) TILs cultured from tumor fragments. Growth medium: RPMI 1640, 100 U/mL penicillin, 100 μg/mL streptomycin, 2 mmol/L *L*-glutamine, 10% human serum, IL-2 (6,000 IU/mL). 2) TILs from enzymatically digested tumor fragments. Solution for enzymatic digestion: collagenase, hyaluronidase, and DNase in RPMI 1640. 18-hr incubation of the fragments on an orbital shaker. Culturing the resulting cell suspension in the TIL growth medium.	1) 69.9 2) 94.1	CD4+ 31.4 (0.3–70), CD8+ 62.4 (37.2–97.6)	N/A	[12]
Breast cancer	Freshly resected material	42 tumors	N/A	100	CD4+ 55.6, (9.1–94.0)	N/A	[24]
Colorectal cancer, stomach cancer	Freshly resected material	33 colorectal tumors, 8 stomach tumors	TILs from enzymatically digested tumor fragments. Four-week culturing of tumor fragments.	Colorectal cancer, 64 Stomach cancer, 43	N/A	N/A	[25]
Colorectal cancer	Fresh material	12 tumors	TILs cultured from tumor fragments. Growth medium: CellGenix GMP DC, 10% human serum, 1% solution of an antibiotic antifungal agent, IL-2 (1,000 IU/mL). 10 ng/mL IL-12 was added when transferring the TIL culture to a perfusion bioreactor for large-scale expansion.	100	First phase of expansion: CD4+ 28.8 (0.6–55.3), CD8+ 64.6 (325–84.5). During the second phase of expansion: CD8+ 85, CD4+ 12.4 (1.7–40.5).	N/A	[26]
Uveal melanoma	Freshly resected material	30 tumors	1) TILs cultured from tumor fragments (22 tumors). Culturing 1–4 tumor fragments in 2 mL of the TIL growth medium containing IL-2. Haft of the growth medium was replaced with fresh one every 2–3 days. 2) TILs from enzymatically digested tumor fragments (12 tumors). Solution for enzymatic digestion: collagenase D 10 mg/mL, DNase I 3 mg/mL. 30-min incubation on a GentleMACS dissociator. 3) TILs from enzymatically digested tumor fragments with additional positive CD3 selection using Dynabeads magnetic beads (25 tumors).	1) 18 2) 42 3) 68	CD4+ 25 (0–91), CD8+ 39 (6–84)	N/A	[27]
Soft tissue sarcoma	Freshly resected material	64 tumors	1) TILs cultured from tumor fragments. Culturing one tumor fragment per well of a 24-well plate in the growth medium containing IL-2 (6,000 IU/mL). 2) TILs from enzymatically digested tumor fragments. Solution for enzymatic digestion: DNase IV (30,000 U/L), hyaluronidase V (100 mg/L), collagenase IV (1,000 mg/L), gentamicin (500 mg/L), penicillin-streptomycin (5,000 U/mL), *L*-glutamine (292 mg/L), amphotericin B (62.5 μg/L). Culturing cell suspension in the growth medium containing IL-2 (6,000 IU/mL).	91	CD8+ 54.2 (3–95.4), CD4+ 2.5 (0.03–44.73)	1.60	[28]

Nosological entity	State of the initial tumor tissue sample	Number of tumor tissue samples in the study	Features of TIL culture generation	Effectiveness of TIL culture generation, %	Percentage of CD4+, CD8+ T- cell populations of the total cell count in the TIL culture, %	Percentage of contaminated TIL cultures	Reference
Primary pancreatic cancer and metastatic melanoma	Freshly resected and frozen material	31 pancreatic cancer samples and 32 metastatic melanoma samples	TILs cultured from tumor fragments. Growth medium: X-Vivo 15 or RPMI 1640, antibiotic, 2% human serum albumin (HSA) or 10% autologous serum.	80	N/A	N/A	[19]
Squamous cell carcinoma of the head and neck	Freshly resected material	47 samples	TILs cultured from tumor fragments. Growth medium: RPMI 1640, 10% fetal bovine serum, 1× ZellShield, 400 μg/mL gentamicin, 50 nM 2-mercaptoethanol, IL-2 (1,000 IU/mL).	36.20	In TILs, CD4+ T cells were the predominant subpopulation in 9 cases (69.2) and CD8+ cells were the predominant subpopulation in 4 cases (30.7)	27.60	[29]
Metastatic non-small cell lung cancer	Freshly resected material	34 samples	TILs cultured from tumor fragments. Duration, 22 days; GMP compliance. Growth medium: IL-2, OKT3 and feeder cells.	82.3	CD4+ 44.5 (6–88), CD8+ 46 (1–92)	N/A	[30]
10 different solid cancers	Freshly resected material	25 samples	TILs cultured from tumor fragments. Growth medium: RPMI 1640, 10% fetal bovine serum, 1% penicillin/streptomycin and 0.5% Fungizone, IL-2 (6,000 IU/mL).	97	CD4+ 59.1 (2–95.8), CD8+ 32 (2.5–65.6). (CD8+ T cells were the predominant subpopulation in 10 patients (> 50%), whereas the CD4+ were the predominant subpopulation in 14 patients).	N/A	[31]
Metastatic colorectal cancer of the liver	Freshly resected material	15 samples	TILs cultured from tumor fragments	66.7	CD4+ 69.4, CD8+ 32.5	N/A	[32]
Metastatic melanoma	Freshly resected material	1135 samples	1) TILs cultured from tumor fragments. Growth medium: 6,000 IU/mL IL-2. 2) TILs cultured from tumor fragments. Growth medium: 30 ng/mL OKT3, 10 μg/mL agonistic anti-4-1BB antibody and 6,000 IU/mL IL-2.	50–80 (mean value, 62 within 11 years; 68 within the past 5 years)	1) CD8+ 74.7 ± 28.1, 2) CD8+ 84.9 ± 9.6	N/A	[33]
Glioblastoma, lung cancer, metastatic colorectal cancer, melanoma metastases of the liver	Freshly resected material	Glioblastoma (7 patients), lung cancer (2 patients), metastatic colorectal cancer (2 patients)	TILs cultured from tumor fragments	100	CD8+ adenocarcinoma (16–36), glioblastoma (2–20)	N/A	[34]

Developing vectors for the *in vivo* delivery of genes to modifying tumor-specific T cells is the last aspect of gene engineering discussed in this review. The technique aims to optimize TIL therapy. Current research in designing viral vectors for *in vivo* gene delivery focuses on restricting viral tropism to specific T-cell markers such as CD3, CD8, CD4, CD62L, and CD5 [53–55]. Thus, the efficacy of retroviruses targeting the peptide–MHC complex (pMHC) for delivering genes, including interleukin-12, to antigen-specific T cells and promoting their *in vivo* expansion, was evaluated in a recently published preprint. Preliminary results of mouse experiments demonstrate that pMHC-targeted viruses are effective vectors for the reprogramming and expansion of tumor-infiltrating T lymphocyte populations *in vivo*.

The reviewed studies demonstrate that diverse approaches are being pursued to optimize the production of T cell-based products, which will broaden the range of their clinical applications [56].

CRYOPRESERVATION IN THE GENERATION OF TUMOR-INFILTRATING T LYMPHOCYTES

The previously mentioned cryopreservation of TILs is highly desirable; in certain cases, it is essential both for manufacturing and in cases when TILs need to be reinfused back into the patient after some time. Cryopreservation implies the slow freezing of cellular products at a rate of ~ 1°C per min in a growth medium containing cryoprotectants, dimethyl sulfoxide (DMSO) being the most commonly used, followed by storage in liquid nitrogen until the product is needed. However, cryopreservation adversely affects all cellular products, including TILs, altering the cytokine production, cytotoxic activity, proliferation, and cell viability [17].

Meanwhile, the therapeutic efficacy of cellular products is directly dependent on the ability of the cells to restore their viability and functionality following thawing.

Although current FDA protocols for both TIL therapy [57] and CAR T therapy [58] permit the use of both fresh cellular products and cryopreserved ones, research into the activity of T cell-based products is ongoing, since the post-thaw viability and functionality of T cells is far from ideal. Importantly, unlike for CAR T cells, the proportion of antigen-specific T cells within the T cell-based product is relatively low, ranging from 0.1 to 9% [59]. Therefore, any reduction in the number of viable cells following the freeze–thaw cycle can critically affect the quality of the T cell-based product. Because the TIL therapy is such a novel technique, very little data on the effects of cryopreservation on TIL quality is available. Three patents have been approved so far. They focus on the optimization of TIL cryopreservation [60–62]. TILs cryopreserved after the pre-REP stage have also been used to produce cellular products in the phase I clinical trial NCT03215810 to assess the TIL therapy in patients with lung cancer [63]. Additionally, as mentioned previously, the approved drug lifileucel is supplied in cryopreserved form, in accordance with the manufacturer's recommendations.

Data on the impact of cryopreservation on CAR T-cell therapy, which has been in clinical use for an appreciably long time, appear somewhat scattered. Based on the information of some CAR T-cell manufacturers, the post-thaw viability ranges from 47.2 to 68.9% [64]. Conversely, another research group has reported an average viability of 97 ± 17.4% in previously cryopreserved CAR T-cell fractions. A total of 79 ready-to-use CAR T infusion products where CAR T cells were expanded to a median value of ~ 1×10^6 cells per kg of body weight (range, 1×10^5 to 1×10^7 cells/kg) were analyzed. The median cryopreservation duration was nine days (range, 1–408 days). Despite the high survival rates in this case, the thawed CAR T cells exhibited increased expression of early apoptotic markers [65]. Another study demonstrated that cryopreservation during the expansion phase does not hinder cell proliferation post-thaw; CAR T cells continued to divide in 86% of cases [66]. Additionally, the study that examined the stability of cryopreserved CAR/TCR T-cell controls showed that these cells remained stable for at least one year after thawing. After 12 months, the viability of thawed cells stood at approximately 80%, remaining stable for at least six hours post-thaw [67].

In an assessment of the tolerance of peripheral blood lymphocytes to cryopreservation following large-scale expansion in the presence of high-dose IL-2, the T cells immediately lost their ability to respond to nonspecific stimulation with phytohemagglutinin after thawing. However, their reactivity was restored within 48 h. Cell viability remained high (> 80%) throughout this process, although each subsequent cryopreservation cycle resulted in a loss of approximately 10–15% of the cells [68].

Comparative analysis with other types of immune cells indicates that regulatory T cells (Tregs) and NK cells also exhibit poor cryopreservation tolerance. One day post-thaw, the proportion of viable NK cells decreased from 64–91% to ~ 34% [69]. A similar trend was observed for Tregs: the percentage of live cells immediately after thawing ranged from 58 to 75%, declining to 20–48% after 24 h [70].

A potential solution towards improving the viability of T cells after cryopreservation is to directly cryopreserve tumor fragments [71–75]. A recent study focusing on the isolation of tumor-infiltrating T lymphocytes from frozen colorectal cancer tissue fragments demonstrated that the efficiencies of TIL culture generation from individual aliquots of cryopreserved fragments of the same tumor were similar after thawing and analyses at different time points, thus indicating data reliability. Furthermore, similar CD4+/CD8+ T-cell ratios were observed in TIL cultures derived from both frozen and fresh tumor fragments [76]. A comparative analysis of TIL generation from fresh vs. frozen tumor samples showed that, although initial expansion occurred at a faster pace in fresh tissue, the total number of viable cells equalized approximately after one week of culturing [77]. In an Australian study where fresh and cryopreserved melanoma fragments derived from the same patients had been transported to a laboratory for further TIL expansion for four days, only the cryopreserved fragments ensured a 100% rate of successful culture generation [78]. Furthermore, in one patent, no phenotypic differences between TILs derived from fresh vs. frozen tumor tissues were listed [59]. Hence, the use of cryopreserved tumor fragments is a viable strategy that allows one to preserve the source of TILs for subsequent expansion, thus addressing the logistical challenges related to the transportation of biological material from the hospital where the tumor had been excised to manufacturing sites, including remote ones. However, standardization is needed for cryopreservation of expanded TILs, as well as tumor fragments and possibly new cryopreservation media, which would improve TIL survival and efficiency in generating TIL-based cellular products.

CONCLUSIONS

Immunotherapy that utilizes tumor-infiltrating T lymphocytes shows great potential as relates to the treatment of various types of cancer. Characterized by a unique specificity to tumor-associated antigens, TILs can effectively destroy malignant cells, especially in melanoma, where this therapy has already proven to be effective.

Despite the encouraging preliminary results, TIL-based therapy is still in its infancy. Some unresolved issues related to therapeutic effectiveness across different tumor types persist, and there exists no standardized protocol for the isolation, expansion, and cryopreservation of TILs. In order to improve therapeutic effectiveness, research aiming to develop unified protocols and optimize the processes related to current challenges is needed.

An important area of focus is exploring novel strategies to augment the antitumor immune response that would be specifically aimed at overcoming the immunosuppressive microenvironment within tumors. Achieving these goals will encourage broader application of TIL-based therapy and improve prognosis for patients with various cancers. ●

This work was conducted under State Assignment Project "T cells" (Research and Technological Development project No. 123032900030-7).

REFERENCES

1. Lee S., Margolin K. // Curr. Oncol. Rep. 2012. V. 14. № 5. P. 468–474. doi: 10.1007/s11912-012-0257-5.
2. Mitchison N.A. // J. Exp. Med. 1955. V. 102. № 2. P. 157–177. doi: 10.1084/jem.102.2.157.
3. Tang Y., Zhang A.X.J., Chen G., Wu Y., Gu W. // Mol. Ther. Oncolytics. 2021. V. 22. P. 410–430. doi: 10.1016/j.omto.2021.07.006.
4. Rosenberg S.A., Spiess P., Lafreniere R. // Science. 1986. V. 233. № 4770. P. 1318–1321. doi: 10.1126/science.3489291.
5. Rosenberg S.A., Packard B.S., Aebersold P.M., Solomon D., Topalian S.L., Toy S.T., Simon P., Lotze M.T., Yang J.C., Seipp C.A., et al. // New Engl. J. Med. 1988. V. 319. № 25. P. 1676–1680. doi: 10.4049/jimmunol.1490019.
6. Search for: TIL, Active, not recruiting studies | Card Results | ClinicalTrials.gov [Electronic resource]. URL:

https://clinicaltrials.gov/search?intr=TIL&aggFilters=status:act (accessed: 13.11.2024).
7. Qiu X., Li S., Fan T., Zhang Y., Wang B., Zhang B., Zhang M., Zhang L. // Discover Oncol. 2024. V. 15. № 1. P. 1–18. doi: 10.1007/s12672-024-01410-5.
8. FDA grants accelerated approval to lifileucel for unresectable or metastatic melanoma | FDA [Electronic resource]. URL: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-grants-accelerated-approval-lifileucel-unresectable-or-metastatic-melanoma (accessed: 13.11.2024).
9. Matsueda S., Chen L., Li H., Yao H., Yu F. // Cancer Immunol., Immunother. 2024. V. 73. № 11. P. 1–20. doi: 10.1007/s00262-024-03793-4.
10. Zhao Y., Deng J., Rao S., Guo S., Shen J., Du F., Wu X., Chen Y., Li M., Chen M., et al. // Cancers. 2022. V. 14. № 17. P. 4160. doi: 10.3390/cancers14174160.
11. Wu R., Forget M.A., Chacon J., Bernatchez C., Haymaker C., Chen J.Q., Hwu P., Radvanyi L.G. // Cancer J. 2012. V. 18. № 2. P. 160–175. doi: 10.1097/PPO.0b013e31824d4465.
12. Dudley M.E., Wunderlich J.R., Shelton T.E., Even J., Rosenberg S.A. // J. Immunother. 2003. V. 26. № 4. P. 332–342. doi; 10.1097/00002371-200307000-00005.
13. Hall M.L., Liu H., Malafa M., Centeno B., Hodul P.J., Pimiento J., Pilon-Thomas S., Sarnaik A.A. // J. Immunother. Cancer. 2016. V. 4. № 1. P. 61. doi: 10.1186/s40425-016-0164-7.
14. Kongkaew T., Thaiwong R., Tudsamran S., Sae-jung T., Sengprasert P., Vasuratna A., Suppipat K., Reantragoon R. // J. Immunol. Methods. 2022. V. 503. P. 113229. doi: 10.3390/vaccines10030457.
15. Nijhuis E.W.P., v/d Wiel-van Kemenade E., Figdor C.G., van Lier R.A.W. // Cancer Immunol. Immunother. 1990. V. 32. № 4. P. 245–250. doi: 10.1007/BF01741708.
16. Flens M.J., Mulder W.M.C., Bril H., von Blomberg van de Flier M.B.E., Scheper R.J., van Lier R.A.W. // Cancer Immunol. Immunother. 1993. V. 37. № 5. P. 323–328. doi: 10.1007/BF01518455.
17. Wickström S., Lövgren T. // Meth. Mol. Biol. 2019. V. 1913. P. 105–118. doi: 10.1007/978-1-4939-8979-9_7.
18. Poch M., Hall M., Joerger A., Kodumudi K., Beatty M., Innamarato P.P., Bunch B.L., Fishman M.N., Zhang J., Sexton W.J., et al. // Oncoimmunology. 2018. V. 7. № 9. doi: 10.1080/2162402X.2018.1476816.
19. Poschke I.C., Hassel J.C., Rodriguez-Ehrenfried A., Lindner K.A.M., Heras-Murillo I., Appel L.M., Lehmann J., Lovgren T., Wickstrom S.L., Lauenstein C., et al. // Clin. Cancer Res. 2020. V. 26. № 16. P. 4289–4301. doi: 10.1158/1078-0432.CCR-19-3845.
20. Rohaan M.W., van den Berg J.H., Kvistborg P., Haanen J.B.A.G. // J. Immunother. Cancer. 2018. V. 6. № 1. P. 102. doi: 10.1186/s40425-018-0391-1.
21. Turcotte S., Gros A., Hogan K., Tran E., Hinrichs C.S., Wunderlich J.R., Dudley M.E., Rosenberg S.A. // J. Immunol. 2013. V. 191. № 5. P. 2217–2225. doi: 10.4049/jimmunol.1300538.
22. Levine B.L., Miskin J., Wonnacott K., Keir C. // Mol. Ther. Methods Clin. Dev. 2017. V. 4. P. 92–101. doi: 10.1016/j.omtm.2016.12.006.
23. Ghorashian S., Kramer A.M., Onuoha S., Wright G., Bartram J., Richardson R., Albon S.J., Casanovas-Company J., Castro F., Popova B., et al. // Nat. Med. 2019. V. 25. № 9. P. 1408–1414. doi: 10.1038/s41591-019-0549-5.
24. Zacharakis N., Huq L.M., Seitter S.J., Kim S.P., Gartner J.J., Sindiri S., Hill V.K., Li Y.F., Paria B.C., Ray S., et al. // J. Clin. Oncol. 2022. V. 40. № 16. P. 1741–1754. doi: 10.1200/JCO.21.02170.
25. Beatty M., Rodriguez-Valentin M., Hall M., Khambati F., Hall A., Pikor L., Langer T.J., Sennino B., Teer J., Fleming J., et al. // J. Immunother. Cancer. 2023. V. 11. № Suppl 1. P. A395–A395. doi: 10.1136/jitc-2023-SITC2023.0346.
26. Albrecht H.C., Gustavus D., Schwanemann J., Dammermann W., Lippek F., Weylandt K.H., Hoffmeister H., Gretschel S. // Cytotherapy. 2023. V. 25. № 5. P. 537–547. doi: 10.1016/j.jcyt.2023.01.009.
27. Gezgin G., Visser M., Ruano D., Santegoets S.J., de Miranda N.F.C.C., van der Velden P.A., Luyten G.P.M., van der Burg S.H., Verdegaal E.M., Jager M.J. // Ophthalmol. Sci. 2022. V. 2. № 2. P. 100132. doi: 10.3390/biomedicines12081758.
28. Mullinax J.E., Hall M., Beatty M., Weber A.M., Sannasardo Z., Svrdlin T., Hensel J., Bui M., Richards A., Gonzalez R.J., et al. // J. Immunother. 2021. V. 44. № 2. P. 63–70. doi: 10.1097/CJI.0000000000000355.
29. Choi S., Hossain M., Lee H., Baek J., Park H.S., Lim C.L., Han D.Y., Park T., Kim J.H., Gong G., et al. // Cancer Immunol. Immunother. 2024. V. 73. № 6. P. 1–14. doi: 10.1007/s00262-024-03691-9.
30. Schoenfeld A.J., Lee S.M., Doger de Spéville B., Gettinger S.N., Häfliger S., Sukari A., Papa S., Rodríguez-Moreno J.F., Graf Finckenstein F., Fiaz R., et al. // Cancer Discov. 2024. V. 14. № 8. P. 1389–1402. doi: 10.1158/2159-8290.CD-23-1334.
31. Kverneland A.H., Chamberlain C.A., Borch T.H., Nielsen M., Mørk S.K., Kjeldsen J.W., Lorentzen C.L., Jørgensen L.P., Riis L.B., Yde C.W., et al. // J. Immunother. Cancer. 2021. V. 9. № 10. doi: 10.1136/jitc-2021-003499.
32. Baek J., Lee H., Gong G., Lim C.-L., Lee H.J. // J. Immunother. Cancer. 2022. V. 10. № Suppl 2. P. A209–A209. doi: 10.1136/jitc-2022-SITC2022.0196.
33. Tavera R.J., Forget M.A., Kim Y.U., Sakellariou-Thompson D., Creasy C.A., Bhatta A., Fulbright O.J., Ramachandran R., Thorsen S.T., Flores E., et al. // J. Immunother. 2018. V. 41. № 9. P. 399–405. doi: 10.1097/CJI.0000000000000230.
34. Yusubalieva G.M., Maratovna Y.G., Petrichuk S.V., Valentinivna P.C., Krivoshapkin A.L., Leonidovich K.A., Kedrova A.G., Genrihovna K.A., Ivanov Yu.V., Viktorovich I.Y., et al. // Clin. Pract. 2020. V. 11. № 1. P. 49–58. doi: 10.17816/clinpract33974.
35. Hernandez-Chacon J.A., Li Y., Wu R.C., Bernatchez C., Wang Y., Weber J.S., Hwu P., Radvanyi L.G. // J. Immunother. 2011. V. 34. № 3. P. 236–250. doi: 10.1097/CJI.0b013e318209e7ec.
36. Chacon J.A., Sarnaik A.A., Chen J.Q., Creasy C., Kale C., Robinson J., Weber J., Hwu P., Pilon-Thomas S., Radvanyi L. // Clin. Cancer Res. 2015. V. 21. № 3. P. 611–621. doi: 10.1158/1078-0432.CCR-14-1934.
37. Shah P., Forget M.A., Frank M.L., Jiang P., Sakellariou-Thompson D., Federico L., Khairullah R., Neutzler C.A., Wistuba I., Chow C.W.B., et al. // J. Immunother. Cancer. 2022. V. 10. № 2. P. e003082. doi: 10.1136/jitc-2021-003082.
38. Bonavita E., Bromley C.P., Jonsson G., Pelly V.S., Sahoo S., Walwyn-Brown K., Mensurado S., Moeini A., Flanagan E., Bell C.R., et al. // Immunity. 2020. V. 53. № 6. P. 1215–1229.e8. doi: 10.1016/j.immuni.2020.10.020.
39. Morotti M., Grimm A.J., Hope H.C., Arnaud M., Des-

buisson M., Rayroux N., Barras D., Masid M., Murgues B., Chap B.S., et al. // Nature. 2024. V. 629. № 8011. P. 426–434. doi: 10.1038/s41586-024-07352-w.
40. Golubovskaya V., Wu L. // Cancers. 2016. V. 8. № 3. P. 36. doi: 10.3390/cancers8030036.
41. Gust J., Ponce R., Liles W.C., Garden G.A., Turtle C.J. // Front. Immunol. 2020. V. 11. P. 577027. doi: 10.3389/fimmu.2020.577027.
42. Quintarelli C., Orlando D., Boffa I., Guercio M., Polito V.A., Petretto A., Lavarello C., Sinibaldi M., Weber G., Del Bufalo F., et al. // Oncoimmunology. 2018. V. 7. № 6. P. doi: 10.1080/2162402X.2018.1433518.
43. Zhou J., Jin L., Wang F., Zhang Y., Liu B., Zhao T. // Protein Cell. 2019. V. 10. № 10. P. 764–769. doi: 10.1007/s13238-019-0643-y.
44. Simpson-Abelson M.R., Mosychuk C., Frank I., Ritthipichai K., Chartier C. // J. Immunotherapy Cancer. 2017. V. 5. Suppl. 2. P. 87. doi: 10.1186/s40425-017-0288-4.
45. Qiu C., Wang J., Zhu L., Cheng X., Xia B., Jin Y., Qin R., Zhang L.X., Hu H., Yan J., et al. // Front. Bioeng. Biotechnol. 2022. V. 10. P. 1027619. doi: 10.3389/fbioe.2022.1027619.
46. Ni J., Griffiths T., Forget M.-A., Sakellariou-Thompson D., Bernatchez C. // Summer Experience. 2021. V. 3. P. 57. doi: 10.52519/00004.
47. https://www.synthekine.com/news/synthekine-presents-positive-initial-results-from-phase-1a-1b-clinical-trial-of-%CE%B1-%CE%B2-biased-il-2-stk-012-for-treatment-of-advanced-solid-tumors/. (Accessed November 14, 2024).
48. Izar B., Zamarin D., Spigel D.R., Hoimes C.J., McDermott D.F., Sehgal K., Najjar Y.G., Schoenfeld A.J., Garon E.B., Sullivan R.J., et al. // Cancer Res. 2024. V. 84. № 7_Suppl. P. CT183–CT183. doi: 10.1158/1538-7445.AM2024-CT183.
49. Sockolosky J.T., Trotta E., Parisi G., Picton L., Su L.L., Le A.C., Chhabra A., Silveria S.L., George B.M., King I.C., et al. // Science. 2018. V. 359. № 6379. P. 1037–1042. doi: 10.1126/science.aar3246.
50. Koliesnik Z., Totagrande M., Burgess R., Tran K.Q., Bauer M., Jayaraman B., Buffone C., Balasubrahmanyam P., Emmerich J., Chaturvedi D., et al. // J. Immunother. Cancer. 2023. V. 11. Suppl. 1. P. A1160–A1160. doi: 10.1136/jitc-2023-SITC2023.1053
51. Vizcardo R., Huang Y., Fioravanti J., Maeda T., Tamaoki N., Yamazaki Y., Kutlu B., Bahl K., Wang B., Zhong Z., et al. // Cancer Res. 2024. V. 84. № 6_Suppl. P. 6593–6593. doi: 10.1158/1538-7445.AM2024-6593
52. Schoenfeld A.J., Betof Warner A., Chesney J.A., Thomas S.S., Hamid O., In G.K., Shoushtari A.N., Samhouri Y., Hari P.N., Ramsingh G., et al. // J. Clin. Oncol. 2024. V. 42. № 16_Suppl. P. TPS9599–TPS9599. doi: 10.1200/JCO.2024.42.16_suppl.TPS959.
53. Agarwal S., Hanauer J.D.S., Frank A.M., Riechert V., Thalheimer F.B., Buchholz C.J. // Mol. Ther. 2020. V. 28. № 8. P. 1783–1794. doi: 10.1016/j.ymthe.2020.05.005
54. Strebinger D., Frangieh C.J., Friedrich M.J., Faure G., Macrae R.K., Zhang F. // Nat. Commun. 2023. V. 14. № 1. P. 1–18. doi: 10.1038/s41467-023-40788-8.
55. Hamilton J.R., Chen E., Perez B.S., Sandoval Espinoza C.R., Kang M.H., Trinidad M., Ngo W., Doudna J.A. // Nat. Biotechnol. 2024. V. 42. № 11. P. 1684–1692. doi: 10.1038/s41587-023-02085-z.
56. Xu E.J.K., Smith B.E., Alberto W.D.C., Walsh M.J., Lim B., Hoffman M.T., Qiang L., Dong J., Garmilla A., Zhao Q.H., et al. // bioRxiv [Preprint]. 2024. doi: 10.1101/2024.09.18.613594.
57. https://www.fda.gov/vaccines-blood-biologics/approved-blood-products/amtagvi. (Accessed December 17, 2024).
58. https://www.fda.gov/vaccines-blood-biologics/cellular-gene-therapy-products/kymriah. (Accessed December 17, 2024).
59. Lowery F.J., Krishna S., Yossef R., Parikh N.B., Chatani P.D., Zacharakis N., Parkhurst M.R., Levin N., Sindiri S., Sachs A., et al. // Science. 2022. V. 375. № 6583. P. 877–884. doi: 10.1126/science.abl5447.
60. Guest R., McCaffrey J., Devices and methods for isolating tumor infiltrating lymphocytes and uses thereof. WO Patent No. WO 2021/123832 A1. 24 June 2021.
61. Frank I., Lotze M.T. Restimulation of cryopreserved tumor infiltrating lymphocytes. WO Patent No. WO 2018/081473 A1. 03 May 2018.
62. Frank I., Lotze M.T. Restimulation of cryopreserved tumor infiltrating lymphocytes. US Patent No. US 10517894 B2. 31 December 2019.
63. Creelan B.C., Wang C., Teer J.K., Toloza E.M., Yao J., Kim S., Landin A.M., Mullinax J.E., Saller J.J., Saltos A.N., et al. // Nat. Med. 2021. V. 27. № 8. P. 1410. doi: 10.1038/s41591-021-01462-y.
64. Hanley P.J. // Mol. Ther. 2019. V. 27. № 7. P. 1213–1214. doi: 10.1016/j.ymthe.2019.06.001.
65. Panch S.R., Srivastava S.K., Elavia N., McManus A., Liu S., Jin P., Highfill S.L., Li X., Dagur P., Kochenderfer J.N., et al. // Mol. Ther. 2019. V. 27. № 7. P. 1275–1285. doi: 10.1016/j.ymthe.2019.05.015.
66. Brezinger-Dayan K., Itzhaki O., Melnichenko J., Kubi A., Zeltzer L.A., Jacoby E., Avigdor A., Shapira Frommer R., Besser M.J. // Front. Oncol. 2022. doi: 10.3389/fonc.2022.1024362.
67. Cai Y., Prochazkova M., Jiang C., Song H.W., Jin J., Moses L., Gkitsas N., Somerville R.P., Highfill S.L., Panch S., et al. // J. Transl. Med. 2021. V. 19. № 1. P. 1–14. doi: 10.1186/s12967-021-03193-7.
68. Sadeghi A., Ullenhag G., Wagenius G., Tötterman T.H., Eriksson F. // Acta Oncol. (Madr.). 2013. V. 52. № 5. P. 978–986. doi: 10.3109/0284186X.2012.737020.
69. Saultz J.N., Otegbeye F. // Front. Immunol. 2023. V. 14. P. 1304689. doi: 10.3389/fimmu.2023.1304689.
70. Kaiser D., Otto N.M., McCallion O., Hoffmann H., Zarrinrad G., Stein M., Beier C., Matz I., Herschel M., Hester J., et al. // Front. Cell. Dev. Biol. 2021. V. 9. doi: 10.3389/fcell.2021.750286.
71. Kuznetsova D., Petrova T., Sharova E., Astrelina T., Varlamova S., Lazarev V. // Ann. Oncol. 2024. V. 35. P. S226. doi: 10.1016/j.annonc.2024.08.047.
72. Poschke I.C., Hassel J.C., Rodriguez-Ehrenfried A., Lindner K.A.M., Heras-Murillo I., Appel L.M., Lehmann J., Lovgren T., Wickstrom S.L., Lauenstein C., et al. // Clin. Cancer Res. 2020. V. 26. № 16. P. 4289–4301. doi: 10.1158/1078-0432.CCR-19-3845.
73. Donlin L.T., Rao D.A., Wei K., Slowikowski K., McGeachy M.J., Turner J.D., Meednu N., Mizoguchi F., Gutierrez-Arcelus M., Lieb D.J., et al. // Arthritis Res. Ther. 2018. V. 20. № 1. P. 1–15. doi: 10.1186/s13075-018-1631-y.
74. Crookes H., McCaffrey J., Hawkins R., Guest R. // Cytotherapy. 2020. V. 22. № 5. P. S142. doi: 10.1016/j.jcyt.2020.03.284.
75. Kobayashi T., Kumagai S., Doi R., Afonina E., Koyama

S., Nishikawa H. // STAR Protoc. 2022. V. 3. № 3. P. 101557. doi: 10.1016/j.xpro.2022.101557.
76. Liang F., Rezapour A., Falk P., Angenete E., Yrlid U. // Cancers (Basel). 2021. V. 13. № 10. P. 2428. doi: 10.3390/cancers13102428.
77. Veerapathran A., Onimus K. Expansion of tils from cryopreserved tumor samples. WO Patent No. WO2020/061429 A1, 26 March 2020.
78. Onimus K., Wells A., Herman C., Tawashi A., Long G.V., Scolyer R.A., Velickovic R., Saw R.P., Pennington T.E., Menzies A.M., et al. // Transplant. Cell. Ther. 2022. V. 28. № 3. P. S226–S227. doi: 10.1016/S2666-6367(22)00447-X.

Extracellular Vesicles As a Source of Biomarkers for Cancer Diagnosis

L. A. Ovchinnikova*, Y. A. Lomakin
Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Moscow, 117997 Russia
*E-mail: leyla_ovchinnikova@yahoo.com
Received: December 11, 2024; in final form, December 30, 2024
DOI: 10.32607/actanaturae.27591

ABSTRACT **Extracellular vesicles (EVs) are secreted by nearly all mammalian cells and play a major role in intercellular communication via the transport of various active biomolecules. In cancer, pathological EVs contribute to tumor progression by participating in metastasis, angiogenesis, and immune evasion. Recent advancements in EV research have revealed their potential as noninvasive biomarkers. This review addresses the latest advancements in EV isolation and characterization techniques, elucidates the molecular mechanisms underlying EV biogenesis, and examines their functional roles in cancer progression. Furthermore, we discuss emerging strategies that leverage EV profiling and molecular composition analysis, in conjunction with liquid biopsy technologies, offering possible breakthroughs in early cancer diagnosis and treatment monitoring. By synthesizing these insights, this review emphasizes the growing significance of EVs as versatile and powerful diagnostic tools in oncology.**
KEYWORDS **extracellular vesicles, EVs, oncology, exosomes, liquid biopsy.**
ABBREVIATIONS **EV – extracellular vesicle; EV-DNA – EV-associated DNA; EV-RNA – EV-associated RNA; cfDNA – cell-free DNA; HCC – hepatocellular carcinoma; CRC – colorectal cancer; PC – pancreatic cancer; LC – lung cancer; BC – breast cancer; PCa – prostate cancer; circRNA – circular RNA; CAF – cancer-associated fibroblast; CDE – CAF-derived exosome; EE – early endosome; ESCRT – endosomal sorting complex required for transport; IDL – intermediate-density lipoprotein; ILV – intraluminal vesicle; MVB – multivesicular body; piRNA – PIWI-interacting RNA; PSA – prostate-specific antigen.**

INTRODUCTION

Extracellular vesicles (EVs) are spherical lipid bilayer particles that are secreted by all types of cells. EVs are usually classified into exosomes and microvesicles (or ectosomes), based on their origin. However, the diversity of EVs extends beyond this classification. Recent studies have identified many other EV subtypes, such as small ectosomes, apoptotic bodies, migrasomes, large oncosomes, and exophers [1]. In addition, cells can release nonvesicular extracellular nanoparticles, such as supermeres, exomeres, and supramolecular attack particles [2]. To create a unified, standardized classification, the International Society for Extracellular Vesicles (ISEV) has regularly published and updated its MISEV guidelines. These guidelines are an important resource for researchers, because they ensure consistency and accuracy in the characterization of EVs.

Exosomes are the type of EVs that have most widely been studied. They range from 30 to 150 nm in diameter. Exosomes are formed during the release of intraluminal vesicles (ILVs) upon fusion of multivesicular bodies (MVBs) with the plasma membrane, resulting in the secretion of these particles into the extracellular space [3, 4]. While exosomes from normal cells facilitate intercellular communication by transporting various molecules (e.g., proteins, DNA, RNA, lipids), exosomes released by tumor cells are involved in tumor progression, metastasis, angiogenesis, and, in some cases, contribute to chemoresistance [5].

This review analyzes current knowledge about EVs released by tumor cells, the role of EVs in cancer progression, and the potential of EVs as biomarkers.

ISOLATION AND CHARACTERIZATION OF EXTRACELLULAR VESICLES

Efficient isolation of EVs is an important step in their investigation, but it is often a non-trivial challenge. There are many EV purification techniques, each with

its advantages and limitations. However, there is no versatile technique for vesicle isolation; the choice of approach depends on the specific purpose of the research. EVs isolation techniques may be classified as follows: (i) high yield but low purity techniques (polymer precipitation, ultrafiltration); (ii) medium yield and purity techniques (differential ultracentrifugation and size exclusion chromatography); (iii) low yield but high purity techniques (gradient ultracentrifugation, affinity isolation, flow cytometry, and microfluidic approaches) [6]. Often, a combination of these techniques can increase EVs yield and purity [7]. In this case, new techniques for EVs isolation from biological fluids have been under development. One of these approaches, ExoArc, uses a high-throughput inertial microfluidic device that efficiently isolates cell-free plasma for comprehensive RNA and EVs analysis. In conjunction with size exclusion chromatography, this technique affords EVs yields 10-fold higher than those obtained with ultracentrifugation techniques [8].

Various methods are used to characterize EVs. One of the most common approaches is direct visualization of EVs using microscopy; in particular transmission electron microscopy (TEM), scanning electron microscopy (SEM), cryo-electron microscopy (cryo-EM), and atomic force microscopy (AFM). The use of TEM to visualize EVs often results in images of cup-shaped EVs due to sample dehydration, whereas AFM and cryo-EM help preserve the original spherical morphology of EVs, representing their structures more accurately [4]. Another method for characterizing EVs is dynamic light scattering (DLS), which is based on the Brownian motion of dispersed particles. DLS measures the light scattering intensity fluctuations induced by particle motion, which enables one to measure their size distribution. This method is useful for studying the hydrodynamic diameter of EVs and providing information on their size and homogeneity in solution. DLS is widely used for the analysis of EVs in their natural environment [9]. Compared with DLS, nanoparticle trajectory analysis (NTA) enables one to track individual nanoparticles, a tool that is particularly efficient in particle size analysis in complex samples. A significant advantage of NTA is the ability to use fluorescent labels, which allows one to distinguish particles based on their fluorescence signals. Therefore, NTA allows for simultaneous analysis of the sizes of different individual EVs labeled with different fluorescent markers [10]. Although DLS is easier to use and provides faster results, NTA ensures higher accuracy, especially when working with heterogeneous samples. These methods provide insights into the morphology and size of EVs, and investigation of surface molecules is equally important and may help determine the origin of the EVs. Flow cytometry can be used to analyze EV surface markers, but the diameter of EVs is below the detection limit of standard cytometers, and specialized kits are used to overcome these limitations. The mode of action of these kits is based on positive selection using antibodies against EV markers (e.g. CD63, CD81), which are adsorbed on the microparticle's surface. EVs bound to antibodies remain on microparticles and can be detected by standard cytometers. These kits are able to help more accurately characterize different EV subtypes, based on surface marker expression levels, and to evaluate their functional properties.

BIOGENESIS AND MOLECULAR COMPOSITION OF EXTRACELLULAR VESICLES

The biogenesis of two main EVs types – exosomes and ectosomes – encompasses various cellular processes (*Fig. 1*). Exosome biogenesis begins with the formation of early endosomes via invagination of the plasma membrane. These early endosomes can either transport incoming (macro)molecules and supramolecular complexes into intraluminal vesicles (ILVs), which are precursors of exosomes, or transport them back to the plasma membrane. As early endosomes mature, they transform into multivesicular bodies (MVBs) that interact with other organelles, such as the Golgi apparatus, endoplasmic reticulum, mitochondria, and phagosomes. Multivesicular bodies can fuse with the plasma membrane, leading to the secretion of exosomes, or fuse with lysosomes and undergo degradation [11].

There are different pathways of intraluminal vesicles formation within multivesicular bodies. These pathways are divided into ESCRT-dependent and ESCRT-independent ones. Four ESCRT complexes (ESCRT-0, ESCRT-I, ESCRT-II, and ESCRT-III) can interact with the enzymes on the endosomal membrane during exosome biogenesis. The classical ESCRT-dependent pathway involves the recognition of ubiquitinated proteins in the endosomal membrane by ESCRT subcomplexes and VPS4-mediated formation of intraluminal vesicles. An alternative pathway is the syndecan–syntenin–ALIX pathway, where vesicle budding and cargo sorting can occur independently of ESCRT, and VPS4 plays a key role in the final detachment step. The ESCRT-independent pathway uses ceramide, generated from sphingomyelin by nSMase2, that forms lipid raft domains and initiates the maturation of intraluminal vesicles within multivesicular bodies. Thus, the molecular composition of released exosomes depends on the pathways they pass through during their formation. However, there are a number of common proteins typical of the most

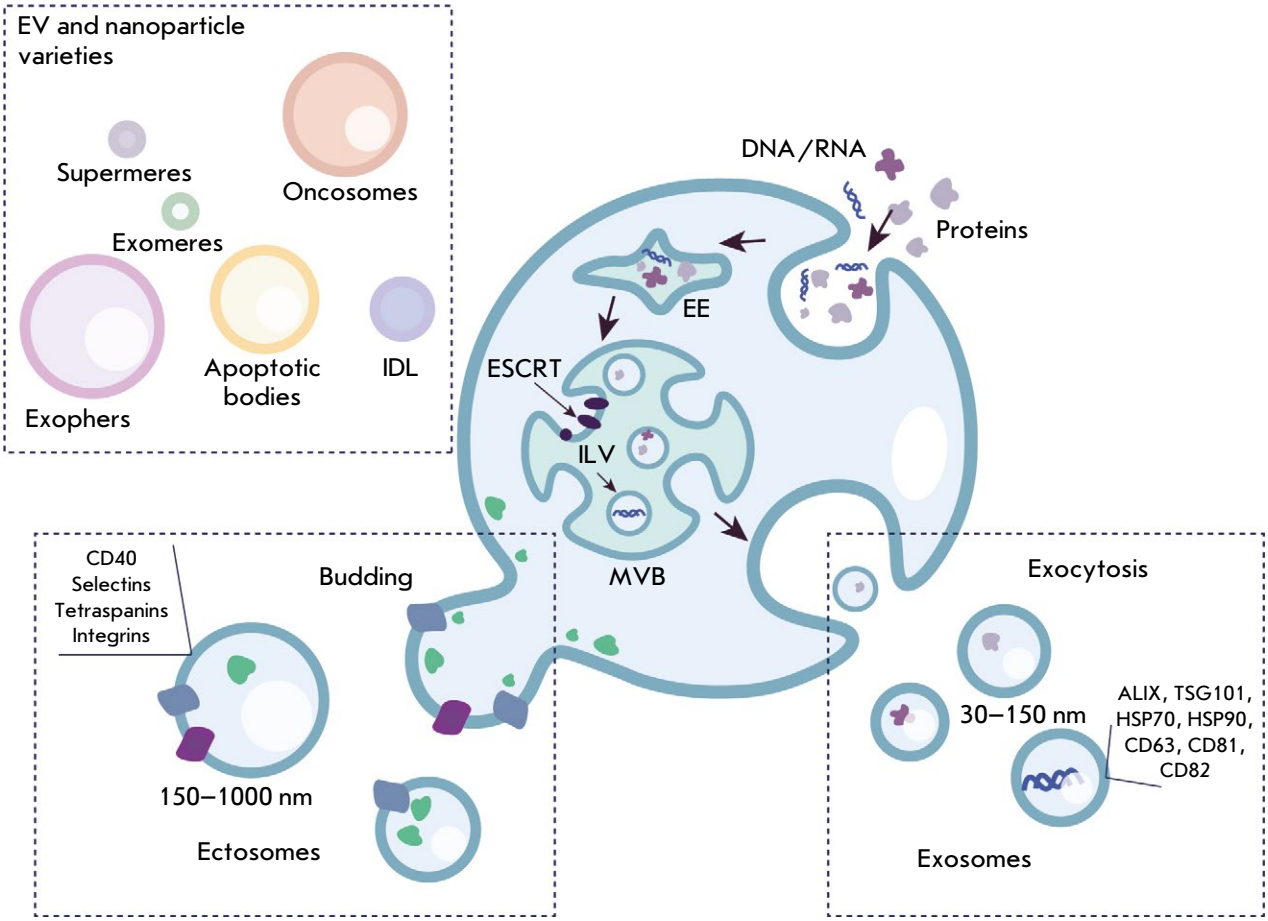


Fig. 1. Schematic of exosomes and microvesicles biogenesis. Exosomes form via the endocytic pathway that starts with the invagination of the plasma membrane and formation of early endosomes (EEs). These endosomes mature into multivesicular bodies (MVBs) containing intraluminal vesicles (ILVs). Following fusion of MVBs with the plasma membrane, ILVs are released as exosomes (30–150 nm) into the extracellular space. Microvesicles are formed by direct budding from the plasma membrane, resulting in larger vesicles (150–1,000 nm). IDL – intermediate-density lipoprotein; ESCRT – endosomal sorting complex required for transport

studied exosomes. These include the proteins involved in membrane transport and fusion (Rab GTPase family and annexins), exosome biogenesis-associated proteins (ESCRT complex proteins, ALIX, TSG101), heat shock proteins (HSP70 and HSP90), tetraspanins (CD63, CD81, and CD82), and cytoskeletal proteins [12]. Besides proteins, characteristic lipids can be found in exosomes. The lipid composition of exosomes depends on the type of producer cells, their developmental stage, and functions. For example, it has been shown that the bis(monoacylglycero) phosphate (BMP) phospholipid stimulates the formation of intraluminal vesicles [13], and that cholesterol is involved in the assembly of the ESCRT system [14]. Sphingomyelin, phospholipids, ganglioside GM3, and cholesterol are the lipids most typical of the exosome membrane [15]. Some exosome membrane lipids may serve as useful diagnostic tools; e.g., phosphatidylserine-exposing exosomes have their origin in malignant cells [16].

Ectosomes (microvesicles), unlike exosomes, bud directly from the plasma membrane of the producer cell (*Fig. 1*). The molecular mechanisms of ectosome biogenesis are less well understood, but the process is known to involve the ESCRT complex and small GTPase proteins such as ARF1, ARF6, and RhoA. These proteins play an important role in the regulation of cytoskeletal dynamics and membrane remodeling [17]. Furthermore, the inward calcium current

and bilayer remodeling play a key role in the formation of ectosomes, influencing their budding from the plasma membrane [18]. Ectosomes carry a wide spectrum of biomolecules, including proteins, lipids, and RNAs, which they transfer to recipient cells, thereby participating in intercellular communication [19]. These EVs rarely possess specific markers, but their association with CD40, selectins, tetraspanins, and integrins has been revealed [20]. In addition, their membranes can incorporate producer cell proteins and lipids [20].

CONTRIBUTION OF EXTRACELLULAR VESICLES TO CANCER PROGRESSION

EVs are secreted by all types of cells and involved in many pathological processes in the human body, including tumor progression. The tumor microenvironment consists of immune and stromal cells, blood vessels, and the extracellular matrix and plays an active role in tumor progression [21]. The interaction between the tumor microenvironment and cancer cells is partially mediated by EVs [22]. EVs and their contents are able to stimulate tumor growth and progression, cause inflammation, and facilitate tumor escape of immune surveillance [23].

One of the main sources of pathogenic cancer cell-derived EVs are cancer-associated fibroblasts (CAFs), which are important components of the tumor microenvironment in solid tumors. These fibroblasts secrete the cytokines and growth factors that play a key role in tumor growth, angiogenesis, inflammation, and metastasis [24]. CAF-derived exosomes (CDEs) are enriched in bioactive molecules, including numerous signaling factors, nucleic acids, functional proteins, and small metabolites, and they likewise play a significant role in tumor microenvironment modulation via the stimulation of tumor growth, metastasis, and resistance to therapy [25]. CDEs have been shown to inhibit mitochondrial oxidative phosphorylation, alter carbon metabolism, and promote tumor growth [26]. These EVs contain metabolites, in particular amino acids, lipids, and citric acid cycle intermediates, that can be utilized by tumor cells [26]. In addition, these EVs enhance the migratory and invasive capabilities of cancer cell lines, such as SKOV-3 and CAOV-3, and they stimulate epithelial–mesenchymal transition, which is largely a product of elevated TGFβ1 levels [27]. In an animal model of breast cancer (BC), CDEs were shown to enhance tumor cell motility and invasive activity [28]. These exosomes were taken up by tumor cells, providing them with Wnt11, a signaling protein associated with tumor progression. In the case of pancreatic cancer, EVs secreted by tumor-associated fibroblasts increased the chemoresistance-inducing factor (Snail) in recipient epithelial cells and promoted their proliferation and capacity for drug resistance. Inhibition of CDE release reduced the survival of co-cultured epithelial cells, signifying the important role of CDEs in maintaining drug resistance [29].

The pathogenic role of tumor-associated fibroblasts and their EVs is well-documented; however, the molecular mechanisms underlying the reprogramming of normal fibroblasts into tumor-associated ones are poorly understood. One potential mechanism involves the EV-mediated transport of pathogenic microRNAs (miRNAs). A new potential pathway of intercellular communication has been identified in melanoma cells inducing fibroblast transformation via EV-transported miRNAs [30]. It has been shown that melanoma cell-secreted EVs deliver miR-92b-3p into normal fibroblasts, and that the accumulation of this miRNA in the cells correlates with their transformation into tumor-associated fibroblasts [29].

Ascites, which is the accumulation of fluid in the peritoneal cavity, often develops in various pathological conditions, including cancers, and it is another component of the tumor microenvironment, as well as an important source of EVs [31]. In high-grade serous ovarian cancer, ascites fluid was shown to contain EVs originating predominantly from macrophages and fibroblasts rather than tumor cells [32]. A proteomic analysis revealed that ascites-specific EV markers were able to predict patient survival more accurately than traditional cellular markers. EVs derived from ascites ($EXO^{Ascites}$) from gastric cancer patients were also shown to stimulate invasiveness and angiogenesis in a three-dimensional autologous tumor spheroid microfluidic system. $EXO^{Ascites}$ delivered the *MET* oncogene into tumor cells, stimulating oncogenic signals. Modified MET-depleted EVs reduced tumor progression, a sign of potential for targeted therapy [33].

EVs play a significant role in the stimulation of tumor angiogenesis. For example, a known angiogenesis inducer, E-cadherin, is secreted in the form of exosomes [34]. In addition, miR-21, which is present in cancer-associated fibroblast EVs, is delivered into endothelial cells in multiple myeloma, where it regulates angiogenesis [35]. EVs also promote the formation of a pre-metastatic niche, a microenvironment meant for the colonization of circulating tumor cells in specific organs. EVs isolated from pancreatic ductal adenocarcinoma were identified as carriers of the migration inhibitory factor (MIF), a key component in the formation of the pre-metastatic niche in the liver. Blocking MIF in these EVs effectively prevented both pre-metastatic niche formation and subsequent liver metastases. These EVs activated hepatic stellate cells and stimulated extracellular matrix remodeling. This

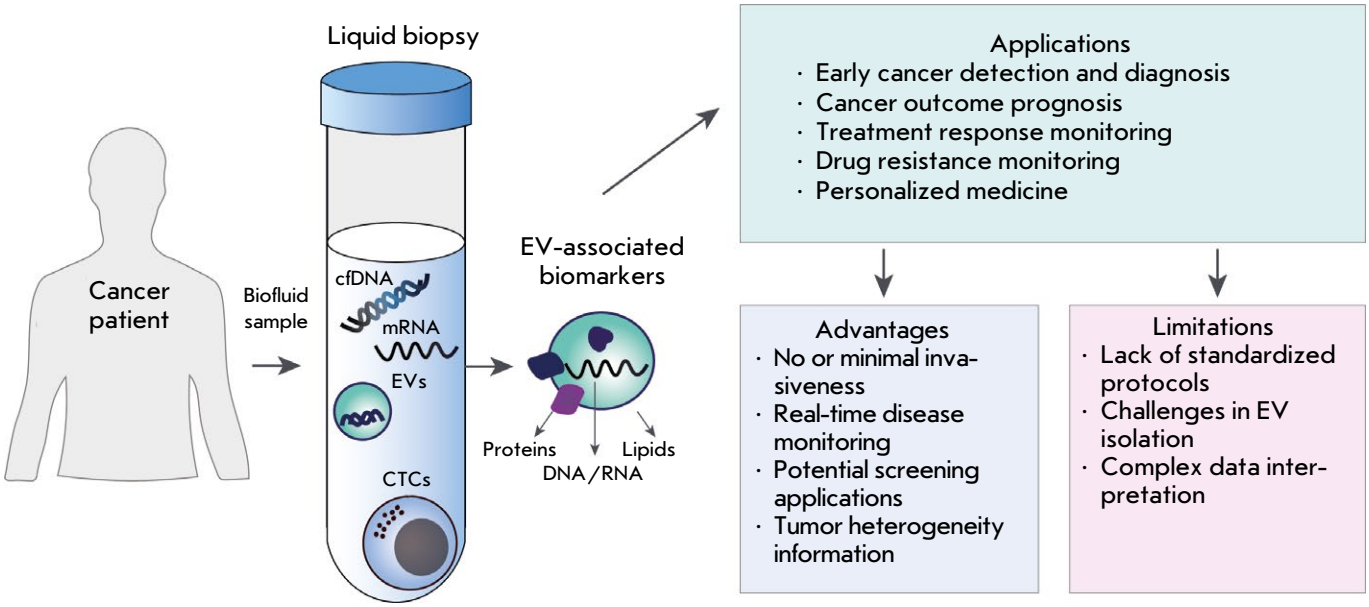


Fig. 2. Application of EVs in liquid biopsy for cancer diagnosis. Key elements analyzed by liquid biopsy include circulating cell-free DNA (cfDNA), extracellular vesicles (EVs), mRNAs, circulating tumor cells (CTCs), and tumor-derived metabolites

process resulted in the accumulation of fibronectin that recruits macrophages, thereby creating a microenvironment supporting liver metastasis [36].

Another input from EVs in tumor progression is their ability to modulate the immune response. EVs isolated from the cells of chronic lymphocytic leukemia patients induced an immunosuppressive phenotype in monocytes. These EVs stimulated the release of CCL2, CCL4, and interleukin-6 and induced PD-L1 expression via delivery of the non-coding RNA hY4 [37]. PD-L1 was also detected on the surface of glioblastoma-derived exosomes that fostered PD-L1-dependent inhibition of T-cell activation [38]. Tumor EVs were shown to transfer fatty acids to dendritic cells, which led to lipid accumulation and increased fatty acid oxidation, causing dendritic cell immune dysfunction [39].

EXTRACELLULAR VESICLES AS A TOOL FOR CANCER DIAGNOSIS. LIQUID BIOPSY

EVs can be isolated from all types of human biological fluids, in particular blood, tears, urine, saliva, cerebrospinal fluid (CSF), etc. This versatility makes EVs a promising tool for cancer diagnosis, especially in terms of liquid biopsy. Liquid biopsy is an innovative technique used to analyze circulating tumor cells, extracellular nucleic acids, and EVs (*Fig. 2*). This minimally invasive method enables real-time monitoring of tumor progression [40]. The advantages of EVs analysis using liquid biopsy are as follows: (1) higher EVs concentrations in biological fluids than in circulating tumor cells; (2) EVs, compared with circulating DNA, provide a better insight into producer cells; and (3) the high biological stability of EVs in the aggressive tumor environment [41]. EVs isolated from tumor cells carry a wide range of cytosolic and surface proteins, DNAs, RNAs, as well as various lipids and glycans; so, they can potentially be used in screening for early cancer stages, monitoring cancers, and predicting the response to therapy. Below, we discuss the application of EV analysis to the diagnosis of the most common cancers using liquid biopsy.

Prostate cancer

Prostate cancer (PCa) is one of the cancers that has been successfully diagnosed using liquid biopsy. Although the introduction of prostate-specific antigen (PSA) testing has significantly improved diagnostics, there remains a need for biomarkers in order to more accurately track disease progression [42]. In a study using plasma from PCa patients, genomic profiling of EV-associated DNA (EV-DNA) provided tumor characteristics and was in correlation with disease progression, whereas the investigation of EV-associated

RNA (EV-RNA) provided insight into the tumor response at early stages of the therapy [43]. Specific miRNAs present in EVs may also be considered as potential biomarkers of PCa. In particular, miR-375, miR-21, and miR-574 were identified in EVs isolated from the serum of PCa patients [44]. In addition, miR-21 and miR-375 were also detected in urinary EVs, indicating that these markers may be used for noninvasive diagnostics [45]. Another EV-associated miRNA, miR-141, was also detected in both the serum and urine of PCa patients, suggesting its potential as a marker for monitoring PCa [46, 47]. It should be noted that PSA was also found in EVs isolated from PCa patients, suggesting that EVs may be used as a source of clinically relevant information [48]. The presence of these specific miRNAs and protein markers in EVs emphasizes their potential role as biomarkers for early detection, progression monitoring, and treatment response assessment in PCa.

Colorectal cancer

Colorectal cancer (CRC) is the third most common malignancy worldwide [49]. Traditional diagnostic methods for CRC are invasive and often painful. The development of new, noninvasive diagnostic tools may reduce mortality rates through earlier diagnosis [50]. Most of the EV-associated biomarkers for CRC are RNAs (in particular, miRNAs). A meta-analysis of 159 publications revealed three miRNAs common to all stages of the disease: miR-146a-5p, miR-22-3p, and miR-23b-3p [51]. In addition, seven miRNAs specific to certain CRC stages were identified: stage I – miR-301a-3p and miR-548i; stage IIIA – miR-23a-3p; and stage IV – miR-194-3p, miR-33a-3p, miR-485-3p, and miR-194-5p [51]. However, the levels of these markers in biological fluids vary significantly, which emphasizes the need for their further validation. Several types of EV-miRNAs have been identified in serum, including let-7a-5p, let-7c-5p, let-7f-5p, let-7d-3p, miR-423-5p, miR-584-5p, miR-30a-5p, miR-99-5p, miR-150-5p, miR-26-5p, and miR-204-5p [52]. A bioinformatics analysis revealed that the let-7 miRNA family targets the key genes in the TGF-β signaling pathway, in particular TGFβRI and SMAD2, which play significant roles in tumorigenesis. In addition, five more EV-miRNAs (hsa-miR-126, hsa-miR-139, hsa-miR-141, hsa-miR-29c, and hsa-miR-423) displaying high potential as CRC markers have been identified. The miRDIP database was used to establish links between these miRNAs and their target mRNAs involved in the regulation of key pathways, such as the B-cell receptor signaling pathway and glycosphingolipid biosynthesis [53]. Long non-coding RNAs (lncRNAs) can also contribute to CRC progression and serve as prognostic markers of the disease [54, 55]. Not only EV-RNAs, but also some proteins present in EVs can be potential markers of the disease. For example, the prion protein PrPC, found in EVs in CRC, is involved in the formation of conditions for metastasis. This occurs due to increased endothelial permeability and the enhanced secretion of angiogenic factors. A potential new therapeutic approach to control CRC metastasis is chemotherapy combined with anti-PrPC therapy [56].

Hepatocellular carcinoma

Hepatocellular carcinoma (HCC) is one of the most common types of primary liver cancer. It's prognosis, despite advances in treatment, remains unfavorable in most cases. Growing evidence suggests that EVs may serve as specific diagnostic – and even prognostic – biomarkers for HCC [57]. MiRNAs stand out among the most studied exosomal biomarkers for HCC. Some exosomal miRNAs can also be used to choose a treatment strategy at late HCC stages [58]. For example, a panel of miRNAs identified as potential biomarkers includes miRNAs overexpressed in HCC patients: miR-224, miR-21, miR-210-3p, miR-93, miR-92b, miR-155, and miR-665 [59]. In contrast, the expression level of miRNAs, such as miR-718, miR-744, miR-9-3p, and miR-125b, is decreased in HCC patients. Combining several miRNAs into diagnostic panels may improve diagnostic accuracy. A combination of miR-26a, miR-29c, and miR-199a was shown to effectively discriminate between HCC patients and healthy subjects (AUC = 0.994), as well as between HCC patients and cirrhosis patients (AUC = 0.965) [60]. RNAs carried by EVs, such as circular RNAs (circRNAs), also demonstrate prognostic potential in HCC. For example, the hsa_circ_0029325 level in EVs may be used to predict disease outcome [61]. Another type of EV-derived RNAs that may be used to diagnose HCC is PIWI-interacting RNAs (piRNAs), which are involved in cancer progression. Expression of serum EV-derived piRNAs is elevated in HCC patients, and some of them (e.g., piR-15254, piR-1029, novel-piR-35395, novel-piR-32132, and novel-piR-43597) are potentially usable in HCC diagnosis even in patients with a low tumor burden [62].

EV proteins may also serve as valuable prognostic biomarkers in HCC. For example, decreased CD31 levels in EVs from HCC patients were shown to correlate with HCC recurrence 12 months after surgery [63]. Proteomic profiling yielded a panel of differentially expressed proteins – VWF, LGALS3BP, TGFB1, SERPINC1, HPX, HP, HBA1, FGA, FGG, and FGB – that may form the basis for an HCC diagnostic panel [64]. MiRNAs, circRNAs, piRNAs, and EV proteins

are promising noninvasive biomarkers for improving HCC diagnosis, prognosis, and treatment monitoring, and this opens up new opportunities for personalized patient care.

Pancreatic cancer

Pancreatic cancer (PC) is the third leading cause of cancer-related deaths [65]. The most common pancreatic cancer is pancreatic ductal adenocarcinoma, which accounts for more than 90% of all PC cases. PC is associated with high mortality; only 10% of patients survive 5 years [66]. Early diagnosis is crucial to improve the prognosis in this disease. Recent advances in machine learning have facilitated the identification of novel potential EV-based biomarkers that may aid in the early diagnosis of PC. Machine learning analysis of EV proteins proposed a panel of seven potential PC biomarkers (mucin-1, sialylated Lewis x antigen, ferritin, fibroblast growth factor 2, human epidermal growth factor 3, leptin, and prolactin, AUC = 0.971) [67]. Another promising PC biomarker, whose concentration is increased in EVs, is glypican-1. Detection of glypican-1 in EVs demonstrated 100% sensitivity and specificity in the diagnosis of all stages of PC, efficiently distinguishing pancreatic cancer patients from healthy subjects or chronic pancreatitis patients (AUC = 1.0) [68]. In addition, miR-21 found in the EVs of PC patients may also be used as a biomarker and prognostic factor of overall survival. Elevated miR-21 levels, in combination with miR-4525 and miR-451a, were shown to exhibit a high potential as biomarkers for the identification of patients with a high recurrence risk and poor prognosis [69]. Elevated miR-191 levels were also detected in a subset of PC patients compared to the controls [70]. Some EV glycans and lipids also appear to have potential as diagnostic tools for PC, emphasizing the significance of diverse EV molecules in the liquid biopsy of this cancer type [71].

Lung cancer

Lung cancer (LC), which affects millions annually, remains one of the most frequently diagnosed cancers and the leading cause of cancer-related mortality [72]. Recent advances in multiplexed EV profiling and machine learning have opened up new opportunities for the study of EVs released by lung cancer cells [73]. For example, a system for detecting EV membrane proteins has been developed based on Forster resonance energy transfer. This system was used to identify potential diagnostic markers for early-stage LC (CEA, PD-L1, EpCAM, and CA125) [74]. Another method based on a dielectrophoretic chip revealed elevated miR-21, miR-191, and miR-192 levels in EVs isolated from the blood plasma of lung cancer patients [75]. Additional EV miRNA panels demonstrated their efficiency in the diagnosis of various LC subtypes at early stages. For example, miR-483-3p was proposed as a biomarker for early small cell lung cancer, and miR-152-3p and miR-1277-5p were proposed for early non-small cell lung cancer [76]. In addition, EVs glycan profiling may also be used in the diagnosis of lung cancer. An EV–GLYPH assay, which is based on microfluidic approaches, was used to identify unique glycan signatures of EVs from non-transformed and malignantly transformed lung cells. In a clinical study, that assay successfully differentiated patients with early-stage lung cancer from those with benign nodules [77].

Breast cancer

Breast cancer (BC) is the most common cancer in women. In high-income countries, breast cancer is estimated to be diagnosed in every eighth woman by age 85 years [78]. Molecular profiling of the EVs in BC is a powerful toll for early noninvasive diagnosis, prognosis, and disease monitoring [79]. Proteomic profiling of EVs isolated from BC cell lines was shown to differentiate between different BC subtypes more effectively than profiling of the tumor cells themselves [80]. It was also noted that the protein composition of EVs secreted by BC cells largely reflects their molecular subtype (e.g., HER2-positive or triple-negative BC) [80]. In another study, the analysis of EVs from the plasma of BC donors identified 10 candidate biomarkers, whose levels were higher in BC patients than in healthy subjects (CD3, CD56, CD2, CD25, CD9, CD44, CD326, CD133/1, CD142, and CD14). The lipid profile of EVs, in particular sphingolipids and phospholipids, was shown to significantly differ from that of the tumor cells secreting EVs, which were more enriched in triglycerides and fatty acids. EVs isolated from the plasma of BC patients are characterized as sources of lipid biomarkers for the early detection of BC and its subtypes (ER/PR+, HER2+, and triple-negative BC) [81]. In addition, miRNAs obtained from EVs may also be used for BC diagnosis [82].

The main markers mentioned in this review are listed in *Table 1*.

INNOVATIVE METHODS FOR IMPROVING EXTRACELLULAR VESICLES DETECTION

An efficient search for EV-based biomarkers requires one to increase the sensitivity of the means used to detect those markers compared with that offered by existing classical methods such as mass spectrometry and Western blotting. The use of artificial intelligence and machine learning methods may significantly improve the detection limit of EV-based

Table 1. EV-associated markers for cancer diagnosis

Biomarker type	Name	Associated cancer	Reference
RNA	miR-21↑	PCa	[44, 45]
		HCC	[59]
		PC	[69, 70]
		LC	[75]
		BC	[82]
	miR-141↑	PCa	[46, 47]
	miR-146a-5p ↑ miR-22-3p ↑ miR-23b-3p ↑ miR-301a-3p ↑ miR-548i ↑ miR-23a-3p ↑ miR-194-3p ↑ miR-33a-3p ↑ miR-485-3p ↑ miR-194-5p ↑	CRC	[51]
	let-7a-5p ↑ let-7c-5p ↑ let-7f-5p ↑ let-7d-3p ↑ miR-423-5p ↑ miR-584-5p ↑ miR-30a-5p ↑ miR-99-5p ↑ miR-150-5p ↑ miR-26-5p ↑ miR-204-5p ↑	CRC	[52]
	miR-126 ↑ miR-139 ↑ miR-141 ↑ miR-29c ↑ miR-423 ↑	CRC	[53]
	miR-224↑ miR-21↑ miR-210-3p ↑ miR-93 ↑ miR-92b ↑ miR-155 ↑ miR-665 ↑	HCC	[59]
	miR-718 ↓ miR-744 ↓ miR-9-3p ↓ miR-125b ↓	HCC	[59]
	miR-26a ↑ miR-29c ↑ miR-199a ↑	HCC	[60]
	hsa_circ_0029325 ↑	HCC	[61]

Biomarker type	Name	Associated cancer	Reference
	piR-15254 ↑ piR-1029 ↑ novel-piR-35395 ↑ novel-piR-32132 ↑ novel-piR-43597 ↑	HCC	[62]
	miR-4525 ↑ miR-451a ↑	PC	[69]
	miR-191 ↑ miR-192 ↑	LC	[75]
	miR-483-3p ↑ miR-152-3p ↑ miR-1277-5p ↑	LC	[76]
	miR-375 ↑	PCa	[44, 45]
	miR-574 ↑	PCa	[44]
Proteins	Cellular prion protein	CRC	[56]
	CD31	HCC	[63]
	Von Willebrand factor Galectin-3-binding protein Transforming growth factor beta 1 Antithrombin III Hemopexin Haptoglobin Hemoglobin subunit alpha 1 Fibrinogen alpha chain Fibrinogen gamma chain Fibrinogen beta chain	HCC	[64]
	Mucin-1 Sialylated Lewis x antigen Ferritin Fibroblast growth factor 2 Epidermal growth factor 3 Leptin Prolactin	PC	[67]
	Glypican-1	PC	[68]
	CEA PD-L1 EpCAM CA125	LC	[74]
	PSA	PCa	[48]
Lipids/ phospholipids	Ceramides Sphingomyelins Hexosylceramides Lysophosphatidylcholines Lysophosphatidylethanolamines Phosphatidylcholines Plasmalogens – phosphatidylethanolamines with an ether bond	BC	[81]

Note. EV – extracellular vesicle; CRC – colorectal cancer; PC – pancreatic cancer; LC – lung cancer; BC – breast cancer; HCC – hepatocellular carcinoma; PCa – prostate cancer; CA125 – cancer antigen 125; CEA – carcinoembryonic antigen; EpCAM – epithelial cell adhesion molecule; PD-L1 – programmed cell death receptor 1 ligand; PSA – prostate-specific antigen.
The up (↑) and down (↓) arrows indicate an increase or a decrease, respectively, in the RNA content in extracellular vesicles in samples from cancer patients compared with those from healthy donors.

biomarkers by liquid biopsy. One of the approaches that improves EV detection is fluorescence polarization using aptamers for the detection of extracellular nanovesicles (FluoPADE) [83]. This method is based on the use of DNA aptamers and fluorescence polarization to detect EVs in human plasma and the culture medium. The specificity of the assay is achieved by fixation of the EVs with antibodies and subsequent detection using a DNA aptamer that targets a specific EV biomarker. This method can be used for early cancer detection, detection of micrometastases, and the monitoring of minimal residual disease. Another approach involves DNA-based barcoding of EVs to explore the protein composition of their surface [84]. One of the advantages of this technology is the ability it affords to investigate the composition of individual exosomes. Also, a method based on nanostructured 3D sensors was developed for the molecular and functional profiling of EVs from cancer stem cells. These highly sensitive sensors were able to detect up to 10 individual EVs in 10 μL, and when combined with artificial intelligence algorithms, allowed one to separate cancer samples from normal ones with 100% sensitivity and 100% specificity [85]. Another method, DNA cascade reaction-triggered individual EV nanoencapsulation (DCR-IEVN), enables the encapsulation of EV subpopulations directly from clinical serum samples. This approach, when integrated with machine learning algorithms, proved highly accurate in diagnostics for HCC [86]. Hoshino et al. performed large-scale proteomic analyses of EVs from various tissues, cells, and biological fluids [87]. They showed that classic EV markers such as CD63, TSG101, flotillins, and ALIX were underrepresented in human plasma EVs. Instead, alternative markers for EV isolation such as MSN, FLNA, STOM, and RAP1B were proposed by the group. Then, machine learning methods were used to identify a panel of EVs proteins specific to certain tumor types. The technique that can be used to classify cancers of unknown primary origin. Proteins and the specific RNAs in individual EVs can be detected using a SPIRFISH technique that combines interferometric reflectance sensor technology with fluorescence *in situ* hybridization, which ensures high detection sensitivity and specificity [88].

Modern EVs research actively uses artificial intelligence. For example, deep learning algorithms were used in miRNA profiling at the individual EV level [89]. This method combines total internal reflection fluorescence (TIRF) imaging, which simultaneously detects several miRNAs in individual EVs, with an algorithm for automated image analysis. Another deep learning algorithm uses nanoplasmonic spectra to analyze mutated exosomal proteins. The technique may be promising in the efforts to monitor the efficiency of cancer therapy [90].

The limited availability of some biological fluids has prompted researchers to develop innovative methods for EVs isolation. It has been proposed to use cellulose nanosheets that can efficiently capture EVs from a small volume of liquid for subsequent sequencing of small RNAs [91]. Liquid biopsy of EVs offers many advantages compared with classical diagnostic methods. First, it is a noninvasive method that can minimize the need for procedures such as puncture or tissue biopsy, providing patients with more options and helping monitor disease progression and therapy effectiveness. Another of the advantages of this method is the ability it affords one to analyze all biological fluids, which allows for a comprehensive characterization of various tumors.

CONCLUSION

EVs are critically involved in tumor progression. The ability to transport biologically active molecules and alter the tumor's microenvironment makes EVs potent mediators of tumor progression, metastasis, and immune evasion. Furthermore, EVs are promising tools in the early diagnosis and monitoring of cancers using liquid biopsy techniques. Recent advances in EV isolation and characterization have significantly improved accuracy and efficiency in their investigation, in particular in the field of oncology. The development of innovative methods such as high-throughput microfluidic platforms and machine learning algorithms has increased capabilities in EV detection and analysis and helped to more thoroughly characterize their molecular composition and functional properties. Therefore, investigation of the abnormalities in the molecular composition of EVs in cancers opens up enormous potential for future personalized medicine and tumor diagnosis. ●

This study was funded by Russian Science Foundation Grant No. 22-14-00219.

REFERENCES

1. Buzas E.I. // Nat. Rev. Immunol. 2023. V. 23. № 4. P. 236–250. doi: 10.1038/s41577-022-00763-8.
2. Jeppesen D.K., Zhang Q., Franklin J.L., Coffey R.J. // Trends Cell Biol. 2023. V. 33. № 8. P. 667–681. doi: 10.1016/j.tcb.2023.01.002.
3. Doyle L.M., Wang M.Z. // Cells. 2019. V. 8. № 7. P. 727. doi: 10.3390/cells8070727.
4. Chuo S.T.-Y., Chien J.C.-Y., Lai C.P.-K. // J. Biomed. Sci. 2018. V. 25. № 1. P. 91. doi: 10.1186/s12929-018-0494-5.
5. Dai J., Su Y., Zhong S., Cong L., Liu B., Yang J., Tao Y., He Z., Chen C., Jiang Y. // Signal Transduct. Target. Ther. 2020. V. 5. № 1. P. 145. doi: 10.1038/s41392-020-00261-0.
6. Clos-Sansalvador M., Monguió-Tortajada M., Roura S., Franquesa M., Borràs F.E. // Eur. J. Cell Biol. 2022. V. 101. № 3. P. 151227. doi: 10.1016/j.ejcb.2022.151227.
7. Ovchinnikova L.A., Tanygina D.Y., Dzhelad S.S., Evtushenko E.G., Bagrov D.V., Gabibov A.G., Lomakin Y.A. // Heliyon. 2024. V. 10. № 24. P. e40940. doi: 10.1016/j.heliyon.2024.e40940.
8. Leong S.Y., Lok W.W., Goh K.Y., Ong H.B., Tay H.M., Su C., Kong F., Upadya M., Wang W., Radnaa E., et al. // ACS Nano. 2024. V. 18. № 8. P. 6623–6637. doi: 10.1021/acsnano.3c12862.
9. Lyu T.S., Ahn Y., Im Y.-J., Kim S.-S., Lee K.-H., Kim J., Choi Y., Lee D., Kang E., Jin G., et al. // PLoS One. 2021. V. 16. № 1. P. e0231994. doi: 10.1371/journal.pone.0231994.
10. Wu S., Zhao Y., Zhang Z., Zuo C., Wu H., Liu Y. // Photonics. 2024. V. 11. № 2. P. 101. doi: 10.3390/photonics11020101.
11. Han Q.-F., Li W.-J., Hu K.-S., Gao J., Zhai W.-L., Yang J.-H., Zhang S.-J. // Mol. Cancer. 2022. V. 21. № 1. P. 207. doi: 10.1186/s12943-022-01671-0.
12. Lee Y.J., Shin K.J., Chae Y.C. // Exp. Mol. Med. 2024. V. 56. № 4. P. 877–889. doi: 10.1038/s12276-024-01209-y.
13. Rabia M., Leuzy V., Soulage C., Durand A., Fourmaux B., Errazuriz-Cerda E., Köffel R., Draeger A., Colosetti P., Jalabert A., et al. // Biochimie. 2020. V. 178. P. 26–38. doi: 10.1016/j.biochi.2020.07.005.
14. Ghadami S., Dellinger K. // Front. Mol. Biosci. 2023. V. 10. P. 1198044. doi: 10.3389/fmolb.2023.1198044.
15. Donoso-Quezada J., Ayala-Mar S., González-Valdez J. // Traffic. 2021. V. 22. № 7. P. 204–220. doi: 10.1111/tra.12803.
16. Lea J., Sharma R., Yang F., Zhu H., Ward E.S., Schroit A.J. // Oncotarget. 2017. V. 8. № 9. P. 14395–14407. doi: 10.18632/oncotarget.14795.
17. Teng F., Fussenegger M. // Adv. Sci. (Weinh). 2020. V. 8. № 1. P. 2003505. doi: 10.1002/advs.202003505.
18. Sun M., Xue X., Li L., Xu D., Li S., Li S.C., Su Q. // Quant. Imaging Med. Surg. 2021. V. 11. № 11. P. 4604–4616. doi: 10.21037/qims-20-1015.
19. Surman M., Stępień E., Hoja-Łukowicz D., Przybyło M. // Clin. Exp. Metastasis. 2017. V. 34. № 3–4. P. 273–289. doi: 10.1007/s10585-017-9844-z.
20. Ratajczak M.Z., Ratajczak J. // Leukemia. 2020. V. 34. № 12. P. 3126–3135. doi: 10.1038/s41375-020-01041-z.
21. Anderson N.M., Simon M.C. // Curr. Biol. 2020. V. 30. № 16. P. R921–R925. doi: 10.1016/j.cub.2020.06.081.
22. Guo Z., Wang G., Yun Z., Li Y., Huang B., Jin Q., Chen Y., Xu L., Lv W. // J. Cancer Res. Clin. Oncol. 2023. V. 149. № 19. P. 17581–17595. doi: 10.1007/s00432-023-05450-2.
23. Paskeh M.D.A., Entezari M., Mirzaei S., Zabolian A., Saleki H., Naghdi M.J., Sabet S., Khoshbakht M.A., Hashemi M., Hushmandi K., et al. // J. Hematol. Oncol. 2022. V. 15. № 1. P. 83. doi: 10.1186/s13045-022-01305-4.
24. Zhou L., Yang K., Andl T., Wickett R.R., Zhang Y. // J. Cancer. 2015. V. 6. № 8. P. 717–726. doi: 10.7150/jca.10865.
25. Peng Z., Tong Z., Ren Z., Ye M., Hu K. // Mol. Med. 2023. V. 29. № 1. P. 66. doi: 10.1186/s10020-023-00665-y.
26. Zhao H., Yang L., Baddour J., Achreja A., Bernard V., Moss T., Marini J.C., Tudawe T., Seviour E.G., San Lucas F.A., et al. // eLife. 2016. V. 5. P. e10250. doi: 10.7554/eLife.10250.
27. Li W., Zhang X., Wang J., Li M., Cao C., Tan J., Ma D., Gao Q. // Oncotarget. 2017. V. 8. № 56. P. 96035–96047. doi: 10.18632/oncotarget.21635.
28. Luga V., Zhang L., Viloria-Petit A.M., Ogunjimi A.A., Inanlou M.R., Chiu E., Buchanan M., Hosein A.N., Basik M., Wrana J.L. // Cell. 2012. V. 151. № 7. P. 1542–1556. doi: 10.1016/j.cell.2012.11.024.
29. Richards K.E., Zeleniak A.E., Fishel M.L., Wu J., Littlepage L.E., Hill R. // Oncogene. 2017. V. 36. № 13. P. 1770–1778. doi: 10.1038/onc.2016.353.
30. Kewitz-Hempel S., Windisch N., Hause G., Müller L., Sunderkötter C., Gerloff D. // J. Extracell. Vesicles. 2024. V. 13. № 9. P. e12509. doi: 10.1002/jev2.12509.
31. Geng Z., Pan X., Xu J., Jia X. // J. Cell Commun. Signal. 2023. V. 17. № 3. P. 391–407. doi: 10.1007/s12079-022-00698-8.
32. Vyhlídalová Kotrbová A., Gömöryová K., Mikulová A., Plešingerová H., Sladeček S., Kravec M., Hrachovinová Š., Potěšil D., Dunsmore G., Blériot C., et al. // J. Extracell. Vesicles. 2024. V. 13. № 3. P. e12420. doi: 10.1002/jev2.12420.
33. Hyung S., Ko J., Heo Y.J., Blum S.M., Kim S.T., Park S.H., Park J.O., Kang W.K., Lim H.Y., Klempner S.J., et al. // Sci. Adv. 2023. V. 9. № 47. P. eadk1098. doi: 10.1126/sciadv.adk1098.
34. Tang M.K.S., Yue P.Y.K., Ip P.P., Huang R.-L., Lai H.-C., Cheung A.N.Y., Tse K.Y., Ngan H.Y.S., Wong A.S.T. // Nat. Commun. 2018. V. 9. № 1. P. 2270. doi: 10.1038/s41467-018-04695-7.
35. Miaomiao S., Xiaoqian W., Yuwei S., Chao C., Chenbo Y., Yinghao L., Yichen H., Jiao S., Kuisheng C. // Sci. Rep. 2023. V. 13. № 1. P. 9671. doi: 10.1038/s41598-023-36092-6.
36. Costa-Silva B., Aiello N.M., Ocean A.J., Singh S., Zhang H., Thakur B.K., Becker A., Hoshino A., Mark M.T., Molina H., et al. // Nat. Cell Biol. 2015. V. 17. № 6. P. 816–826. doi: 10.1038/ncb3169.
37. Haderk F., Schulz R., Iskar M., Cid L.L., Worst T., Willmund K.V., Schulz A., Warnken U., Seiler J., Benner A., et al. // Sci. Immunol. 2017. V. 2. № 13. P. eaah5509 doi: 10.1126/sciimmunol.aah5509.
38. Ricklefs F.L., Alayo Q., Krenzlin H., Mahmoud A.B., Speranza M.C., Nakashima H., Hayes J.L., Lee K., Balaj L., Passaro C., et al. // Sci. Adv. 2018. V. 4. № 3. P. eaar2766. doi: 10.1126/sciadv.aar2766.
39. Yin X., Zeng W., Wu B., Wang L., Wang Z., Tian H., Wang L., Jiang Y., Clay R., Wei X., et al. // Cell Rep. 2020. V. 33. № 3. P. 108278. doi: 10.1016/j.celrep.2020.108278.
40. Poulet G., Massias J., Taly V. // Acta Cytol. 2019. V. 63. № 6. P. 449–455. doi: 10.1159/000499337.
41. Yu D., Li Y., Wang M., Gu J., Xu W., Cai H., Fang X., Zhang X. // Mol. Cancer. 2022. V. 21. № 1. P. 56. doi: 10.1186/s12943-022-01509-9.
42. Ramirez-Garrastacho M., Bajo-Santos C., Line A., Martens-Uzunova E.S., de la Fuente J.M., Moros M., Soekmadji C., Tasken K.A., Llorente A. // Br. J. Cancer. 2022.

V. 126. № 3. P. 331–350. doi: 10.1038/s41416-021-01610-8.
43. Casanova-Salas I., Aguilar D., Cordoba-Terreros S., Agundez L., Brandariz J., Herranz N., Mas A., Gonzalez M., Morales-Barrera R., Sierra A., et al. // Cancer Cell. 2024. V. 42. № 7. P. 1301–1312.e7. doi: 10.1016/j.ccell.2024.06.003.
44. Li M., Rai A.J., DeCastro G.J., Zeringer E., Barta T., Magdaleno S., Setterquist R., Vlassov A.V. // Methods. 2015. V. 87. P. 26–30. doi: 10.1016/j.ymeth.2015.03.009.
45. Foj L., Ferrer F., Serra M., Arévalo A., Gavagnach M., Giménez N., Filella X. // Prostate. 2017. V. 77. № 6. P. 573–583. doi: 10.1002/pros.23295.
46. Li W., Dong Y., Wang K.J., Deng Z., Zhang W., Shen H.F. // Neoplasma. 2020. V. 67. № 6. P. 1314–1318. doi: 10.4149/neo_2020_191130N1234.
47. Li Z., Ma Y.-Y., Wang J., Zeng X.-F., Li R., Kang W., Hao X.-K. // Onco Targets Ther. 2016. V. 9. P. 139–148. doi: 10.2147/OTT.S95565.
48. Logozzi M., Angelini D.F., Iessi E., Mizzoni D., Di Raimo R., Federici C., Lugini L., Borsellino G., Gentilucci A., Pierella F., et al. // Cancer Lett. 2017. V. 403. P. 318–329. doi: 10.1016/j.canlet.2017.06.036.
49. Sung H., Ferlay J., Siegel R.L., Laversanne M., Soerjomataram I., Jemal A., Bray F. // CA Cancer J. Clin. 2021. V. 71. № 3. P. 209–249. doi: 10.3322/caac.21660.
50. Brenne S.S., Madsen P.H., Pedersen I.S., Hveem K., Skorpen F., Krarup H.B., Giskeødegård G.F., Laugsand E.A. // Br. J. Cancer. 2023. V. 129. № 5. P. 861–868. doi: 10.1038/s41416-023-02337-4.
51. Long F., Tian L., Chai Z., Li J., Tang Y., Liu M. // Front. Med. (Lausanne). 2022. V. 9. P. 881788. doi: 10.3389/fmed.2022.881788.
52. Bakhsh T., Alhazmi S., Farsi A., Yusuf A.S., Alharthi A., Qahl S.H., Alghamdi M.A., Alzahrani F.A., Elgaddar O.H., Ibrahim M.A., et al. // Sci. Rep. 2024. V. 14. № 1. P. 8902. doi: 10.1038/s41598-024-58536-3.
53. Ma J., Wang P., Huang L., Qiao J., Li J. // BMC Med. Genomics. 2021. V. 14. № 1. P. 60. doi: 10.1186/s12920-021-00905-2.
54. Hu D., Zhan Y., Zhu K., Bai M., Han J., Si Y., Zhang H., Kong D. // Cell. Physiol. Biochem. 2018. V. 51. № 6. P. 2704–2715. doi: 10.1159/000495961.
55. Yu M., Song X.-G., Zhao Y.-J., Dong X.-H., Niu L.-M., Zhang Z.-J., Shang X.-L., Tang Y.-Y., Song X.-R., Xie L. // Front. Oncol. 2021. V. 11. P. 618967. doi: 10.3389/fonc.2021.618967.
56. Yun C.-W., Lee J.-H., Go G., Jeon J., Yoon S., Lee S.-H. // Cancers (Basel). 2021. V. 13. № 9. P. 2144. doi: 10.3390/cancers13092144.
57. Sasaki R., Kanda T., Yokosuka O., Kato N., Matsuoka S., Moriyama M. // Int. J. Mol. Sci. 2019. V. 20. № 6. P. 1406. doi: 10.3390/ijms20061406.
58. Liu M., Lai Z., Yuan X., Jin Q., Shen H., Rao D., Huang D. // Mol. Med. 2023. V. 29. № 1. P. 136. doi: 10.1186/s10020-023-00731-5.
59. Sorop A., Constantinescu D., Cojocaru F., Dinischiotu A., Cucu D., Dima S.O. // Int. J. Mol. Sci. 2021. V. 22. № 9. P. 4997. doi: 10.3390/ijms22094997.
60. Yang J., Dong W., Zhang H., Zhao H., Zeng Z., Zhang F., Li Q., Duan X., Hu Y., Xiao W. // Front. Cell Dev. Biol. 2022. V. 10. P. 927251. doi: 10.3389/fcell.2022.927251.
61. Yin K.-L., Sun T., Duan Y.-X., Ye W.-T., Ming Li, Liao R. // Discov. Oncol. 2024. V. 15. № 1. P. 212. doi: 10.1007/s12672-024-01060-7.
62. Rui T., Wang K., Xiang A., Guo J., Tang N., Jin X., Lin Y., Liu J., Zhang X. // Int. J. Nanomedicine. 2023. V. 18. P. 1989–2001. doi: 10.2147/IJN.S398462.
63. Juratli M.A., Pollmann N.S., Oppermann E., Mohr A., Roy D., Schnitzbauer A., Michalik S., Vogl T., Stoecklein N.H., Houben P., et al. // Sci. Rep. 2024. V. 14. № 1. P. 5322. doi: 10.1038/s41598-024-55888-8.
64. Zhao L., Shi J., Chang L., Wang Y., Liu S., Li Y., Zhang T., Zuo T., Fu B., Wang G., et al. // ACS Omega. 2021. V. 6. № 1. P. 827–835. doi: 10.1021/acsomega.0c05408.
65. Ilic I., Ilic M. // World J. Gastroenterol. 2022. V. 28. № 32. P. 4698–4715. doi: 10.3748/wjg.v28.i32.4698.
66. Connor A.A., Gallinger S. // Nat. Rev. Cancer. 2022. V. 22. № 3. P. 131–142. doi: 10.1038/s41568-021-00418-1.
67. Hinestrosa J.P., Sears R.C., Dhani H., Lewis J.M., Schroeder G., Balcer H.I., Keith D., Sheppard B.C., Kurzrock R., Billings P.R. // Commun. Med. (London). 2023. V. 3. № 1. P. 146. doi: 10.1038/s43856-023-00351-4.
68. Melo S.A., Luecke L.B., Kahlert C., Fernandez A.F., Gammon S.T., Kaye J., LeBleu V.S., Mittendorf E.A., Weitz J., Rahbari N., et al. // Nature. 2015. V. 523. № 7559. P. 177–182. doi: 10.1038/nature14581.
69. Kawamura S., Iinuma H., Wada K., Takahashi K., Minezaki S., Kainuma M., Shibuya M., Miura F., Sano K. // J. Hepatobiliary Pancreat. Sci. 2019. V. 26. № 2. P. 63–72. doi: 10.1002/jhbp.601.
70. Goto T., Fujiya M., Konishi H., Sasajima J., Fujibayashi S., Hayashi A., Utsumi T., Sato H., Iwama T., Ijiri M., et al. // BMC Cancer. 2018. V. 18. № 1. P. 116. doi: 10.1186/s12885-018-4006-5.
71. Zhang W., Campbell D.H., Walsh B.J., Packer N.H., Liu D., Wang Y. // J. Nanobiotechnology. 2022. V. 20. № 1. P. 446. doi: 10.1186/s12951-022-01641-0.
72. Thai A.A., Solomon B.J., Sequist L.V., Gainor J.F., Heist R.S. // Lancet. 2021. V. 398. № 10299. P. 535–554. doi: 10.1016/S0140-6736(21)00312-3.
73. Huang G., Zheng W., Zhou Y., Wan M., Hu T. // Acta Pharm. Sin. B. 2024. V. 14. № 9. P. 3855–3875. doi: 10.1016/j.apsb.2024.06.010.
74. Xiao S., Yao Y., Liao S., Xu B., Li X., Zhang Y., Zhang L., Chen Q., Tang H., Song Q., et al. // Nano Lett. 2023. V. 23. № 17. P. 8115–8125. doi: 10.1021/acs.nanolett.3c02193.
75. Chen J., Xu Y., Wang X., Liu D., Yang F., Zhu X., Lu Y., Xing W. // Lab. Chip. 2019. V. 19. № 3. P. 432–443. doi: 10.1039/c8lc01193a.
76. Jiang Y.-F., Wei S.-N., Geng N., Qin W.-W., He X., Wang X.-H., Qi Y.-P., Song S., Wang P. // Sci. Rep. 2022. V. 12. № 1. P. 17201. doi: 10.1038/s41598-022-22194-0.
77. Zhou Q., Niu X., Zhang Z., O'Byrne K., Kulasinghe A., Fielding D., Möller A., Wuethrich A., Lobb R.J., Trau M. // Adv. Sci. (Weinh). 2024. V. 11. № 33. P. e2401818. doi: 10.1002/advs.202401818.
78. Britt K.L., Cuzick J., Phillips K.-A. // Nat. Rev. Cancer. 2020. V. 20. № 8. P. 417–436. doi: 10.1038/s41568-020-0266-x.
79. Bandu R., Oh J.W., Kim K.P. // Proteomics. 2024. V. 24. № 11. P. e2300062. doi: 10.1002/pmic.202300062.
80. Rontogianni S., Synadaki E., Li B., Liefaard M.C., Lips E.H., Wesseling J., Wu W., Altelaar M. // Commun. Biol. 2019. V. 2. P. 325. doi: 10.1038/s42003-019-0570-8.
81. Dorado E., Doria M.L., Nagelkerke A., McKenzie J.S., Maneta-Stavrakaki S., Whittaker T.E., Nicholson J.K., Coombes R.C., Stevens M.M., Takats Z. // J. Extracell. Vesicles. 2024. V. 13. № 3. P. e12419. doi: 10.1002/jev2.12419.
82. Lee Y., Ni J., Beretov J., Wasinger V.C., Graham P., Li Y. // Mol. Cancer. 2023. V. 22. № 1. P. 33. doi: 10.1186/

s12943-023-01741-x.
83. Pham C.V., Chowdhury R., Patel S., Jaysawal S.K., Hou Y., Xu H., Jia L., Zhang Y.-M., Wang X., Duan W., et al. // J. Extracell. Vesicles. 2024. V. 13. № 9. P. e12502. doi: 10.1002/jev2.12502.
84. Wu D., Yan J., Shen X., Sun Y., Thulin M., Cai Y., Wik L., Shen Q., Oelrich J., Qian X., et al. // Nat. Commun. 2019. V. 10. № 1. P. 3854. doi: 10.1038/s41467-019-11486-1.
85. Haldavnekar R., Venkatakrishnan K., Tan B. // ACS Nano. 2022. V. 16. № 8. P. 12226–12243. doi: 10.1021/acsnano.2c02971.
86. Li X., Liu Y., Fan Y., Tian G., Shen B., Zhang S., Fu X., He W., Tao X., Ding X., et al. // ACS Nano. 2024. V. 18. № 17. P. 11389–11403. doi: 10.1021/acsnano.4c01310.
87. Hoshino A., Kim H.S., Bojmar L., Gyan K.E., Cioffi M., Hernandez J., Zambirinis C.P., Rodrigues G., Molina H., Heissel S., et al. // Cell. 2020. V. 182. № 4. P. 1044–1061.e18. doi: 10.1016/j.cell.2020.07.009.
88. Troyer Z., Gololobova O., Koppula A., Liao Z., Horns F., Elowitz M.B., Tosar J.P., Batish M., Witwer K.W. // ACS Nano. 2024. V. 18. № 39. P. 26568–26584. doi: 10.1021/acsnano.4c03679.
89. Zhang X.-W., Qi G.-X., Liu M.-X., Yang Y.-F., Wang J.-H., Yu Y.-L., Chen S. // ACS Sens. 2024. V. 9. № 3. P. 1555–1564. doi: 10.1021/acssensors.3c02789.
90. Kim S., Choi B.H., Shin H., Kwon K., Lee S.Y., Yoon H.B., Kim H.K., Choi Y. // ACS Sens. 2023. V. 8. № 6. P. 2391–2400. doi: 10.1021/acssensors.3c00681.
91. Yokoi A., Yoshida K., Koga H., Kitagawa M., Nagao Y., Iida M., Kawaguchi S., Zhang M., Nakayama J., Yamamoto Y., et al. // Nat. Commun. 2023. V. 14. № 1. P. 6915. doi: 10.1038/s41467-023-42593-9.

Identification of Chalcone Synthase Genes from Garlic (*Allium sativum* L.) and Their Expression Levels in Response to Stress Factors

O. K. Anisimova, A. V. Shchennikova, E. Z. Kochieva, M. A. Filyushin*
Skryabin Institute of Bioengineering, Federal Research Centre "Fundamentals of Biotechnology" of the Russian Academy of Sciences, Moscow, 119071 Russia
*E-mail: michel7753@mail.ru
Received: February 19, 2025; in final form, March 13, 2025
DOI: 10.32607/actanaturae.27639

ABSTRACT **A plant's defense response involves the accumulation of flavonoids, whose biosynthetic pathway in garlic *Allium sativum* L. remains not characterized. In this work, we identified eight *AsCHS1–8* genes of chalcone synthases in the *A. sativum* genome which presumably catalyze the first stage of flavonoid synthesis in garlic plants. These genes were found to be localized on 4 chromosomes: *AsCHS2*, *6–8* contain 1 to 2 introns, whereas *AsCHS1*, *3–5* are intronless. The analysis of the organ-specific gene expression profiles revealed significant transcript levels for *AsCHS3* and *8*. Only *AsCHS8* was shown to change its expression level in response to abiotic stressors (salinity, drought, cold) and exogenous phytohormones (abscisic acid, methyl jasmonate). These findings suggest that two out of the eight genes, *AsCHS3* and *8*, control flavonoid synthesis during garlic development, with *AsCHS8* being the most active chalcone synthase gene. The other six genes (*AsCHS1*, *2*, *4–7*) may be involved in flavonoid biosynthesis in highly specialized cells/tissues/organs or the developmental stages of the garlic plant. Our results on the identification and characterization of garlic chalcone synthase genes *AsCHS1–8* may facilitate further analysis of the mechanisms that regulate stress adaptation in *A. sativum* and other *Allium* species.**
KEYWORDS **flavonoid biosynthesis, chalcone synthase CHS, *CHS* gene family, stress response, garlic *Allium sativum* L.**
ABBREVIATIONS **CHS – chalcone synthase; ABA – abscisic acid; MeJA – methyl jasmonate; qPCR – real-time PCR.**

INTRODUCTION

A plant's defense response is associated with the accumulation of flavonoids, a class of plant polyphenols that includes more than 6,900 secondary metabolites displaying a wide range of activities in plant development [1, 2]. Flavonoids, owing to their antioxidant capability [3, 4], play an important role in plants' protection from biotic and abiotic stress factors [5, 6], while also capable of antioxidant, immunomodulatory, antibacterial, and other effects on the human body [7].

The flavonoid biosynthesis pathway is highly conserved. To date, both of the structural (enzyme) genes that control different stages of the biosynthesis and the genes that coordinate the activity of structural genes have been identified in many plant species [8–12]. The key enzymes in the pathway are chalcone synthases (CHS, EC 2.3.1.74), which initiate flavonoid biosynthesis and are structurally conserved in plants [7, 13–15]. In many plant species, *CHS* genes have been shown to be represented in the genome by a family of paralogous copies resulting from evolutionary duplications and mutations of ancestral genes, followed by the functional diversification of paralogs [16–21]. The number of *CHS* family members varies significantly among plant species [10, 22, 23].

One of the economically significant monocot species, garlic *Allium sativum* L. (Amaryllidaceae family, Asparagales order), is not only an important vegetable crop, but is also used in medicine owing to its antioxidant properties [24]. Among other antioxidants, garlic bulbs are rich in flavonoids; in particular, quercetin [24].

The uniqueness of the *A. sativum* species is rooted in its intrinsic asexual reproduction; rare fertile specimens, collected in Central Asia, quickly lose their fertility upon artificial cultivation [25]. New garlic genotypes appear through mutations in vegetative clones, which lead to phenotypic changes [26]. The success in selection is made easier by the high degree of variability of morphophysiological traits, which is characteristic of *A. sativum* [25, 27]; in particular during adaptation to various unfavorable conditions [25], something that is believed to be associated with the evolution of the flavonoid pathway [28].

Thus, the study of flavonoid biosynthesis pathway genes in *A. sativum*, in particular chalcone synthase (*CHS*) family genes, may contribute to our understanding of the regulation of this metabolic pathway, as well as that of the evolution and ontogenetic features of this species. In addition, this will open up new opportunities to characterize world garlic collections and select stress-resistant genotypes with an improved dietary component for plant breeding. *CHS* genes in garlic have not been studied to date. Among other *Allium* species, only onion (*A. cepa*) has been found to possess *CHS-A* and *CHS-B* homologs that are associated with bulb color [29] and to activate *CHS* gene expression in response to fungal infection [30]. The entire *CHS* family (six genes) has been identified only in one of the species that are the most closely related to the genus *Allium*, *Asparagus officinalis* (order Asparagales) [18]. Also, in 2020, the *A. sativum* genome was sequenced and assembled and the transcriptome of certain organs of the garlic plant was sequenced [31], which enables the identification and characterization of gene families.

In this work, we identified and characterized the family of *CHS* genes encoding garlic chalcone synthases and analyzed the expression dynamics of these genes in response to abiotic stressors (drought, salinity, cold) and treatment with phytohormones.

EXPERIMENTAL

Identification and structural characterization of garlic *CHS* genes

Genes were searched in the *A. sativum* cv. Ershuizao genome and transcriptomes (PRJNA606385, Garlic.V2.fa; AlliumDB, https://allium.qau.edu.cn/). *Arabidopsis thaliana* L. chalcone synthases (AT1G02050, AT4G00040, AT4G34850, and AT5G13930) were used as references.

Sequence alignment was performed with MEGA 7.0 (https://www.megasoftware.net/). The exon–intron structures of the *AsCHS* genes were determined by comparing genomic and transcriptomic data (PRJNA606385, Garlic.V2.fa), and the *cis*-regulatory elements in the *AsCHS* gene promoters (2 kbp upstream of the start codon) were identified using PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/). To characterize the AsCHS proteins, the following data were analyzed: the conserved domains and motifs (NCBI-CDD, http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi); Multiple Expectation maximizations for Motif Elicitation (MEME) 5.5.7, http://meme-suite.org/tools/meme; published data [10]); the molecular weight (MW), isoelectric point (pI), and the grand average of the hydrophobicity index (GRAVY) (ExPASy, https://web.expasy.org/protparam/); AsCHS functions (PANNZER, http://ekhidna2.biocenter.helsinki.fi/sanspanz/). The phylogenetic analysis of the chalcone synthases was performed (MEGA 7.0, Neighbor-Joining, bootstrap 1000) by using a comparison of the AsCHS amino acid sequences with those of homologs from *A. thaliana*, *Solanum lycopersicum* L. (tomato), *Capsicum annuum* L. (pepper) (NCBI, https://www.ncbi.nlm.nih.gov/), *A. cepa* (onion), and *A. fistulosum* (Welsh onion) (AlliumDB, https://allium.qau.edu.cn/).

Analysis of the expression pattern of *CHS* genes in different organs of the garlic plant

Expression of the identified *AsCHS* genes in garlic organs was evaluated *in silico*, based on the available transcriptomic data for *A. sativum* cv. Ershuizao [31], and visualized as a heat map (Heatmapper, http://www.heatmapper.ca/expression/). The expression was quantified as FPKM (fragments per kilo base of transcript per million mapped fragments).

The expression pattern of *AsCHS* genes was analyzed by qPCR in the roots, basal plate, bulb, pseudostem, and leaves of garlic plants (cv. Sarmat) grown in open ground in 2024 (Federal Scientific Vegetable Center, Moscow region). The material was ground in liquid nitrogen and used to obtain total RNA (RNeasy Plant Mini Kit, RNase free DNasy set; QIAGEN, Germany) and cDNA (GoScript™ Reverse Transcription System, Promega, USA). The identified *AsCHS* sequences were used to develop specific primers (*Table 1*). *GAPDH* and *UBQ* were used as reference genes [32, 33]. The reaction mixture included 3 ng of cDNA and a SYBR GreenI- and ROX-containing reaction mixture for qPCR (Syntol, Russia). The reaction was conducted using the CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories, USA) in two biological and three technical replicates; the program was as follows: 95°C for 5 min; 40 cycles (95°C for 15 s and 62°C for 50 s). Data were statistically processed (Two-way ANOVA) and visualized in GraphPad Prism v.8 (https://www.graphpad.com).

Table 1. The primer sequences used for the qPCR analysis

Gene	Primer sequence (5'→3')[1]
AsCHS1	F-CGAAGGCCCAGCCACCATT R-CGGTCATGTGCTCGCTGTTG
AsCHS2	F-CACCAACTGCAACAACCTTGAC R-CTCCGGGTATGTGGCCAGT
AsCHS3	F-CAAGACGAATACCCAGACTACTAT R-GATGTCTTCGGACAGGTGCATA
AsCHS4	F-GTACCCAGACTACTACTTCCGT R-ATCTTCGGACAGGTGCATGTAC
AsCHS5	F-GTACCCAGACTACTACTTCCGT R-CAGGTGCATGTAGCGTTTTCTG
AsCHS6	F-CTCTTCTGGATTCCGCATCCT R-CTGCCATTGACCTCTTCCTCA
AsCHS7	F-GCACCGATCTCGCCATGAG R-TAAGCGCTGTTTGATGGTCGG
AsCHS8	F-CTATCGGTACAGCCGTGCCT R-CATGTAGGCCGTCATGTTTGG
GAPDH	F-CCATGTTTGTTGTTGGTGTGAATGAG R-TGGTGCAGCTAGCGTTGGAGAC
UBQ	F-AAGCCAAGATACAGGACAAG R-GCATACCACCTCTCAATCTC

[1]F – forward primer; R – reverse primer

Simulation of stress (drought, salinity, cold, abscisic acid, methyl jasmonate, and darkness) in garlic plants and analysis of the response dynamics of *AsCHS* gene expression

The experiment involved 10-day-old plants (cv. Sarmat) grown in transparent glass cups in water (experimental climate control facility (ECCF), Research Centre of Biotechnology RAS; day/night – 16/8 h, 22/16°C; illumination 190 μM/(m^2·s)); bulb cloves were fixed so that only the root zone was in the water. The experimental plants were placed in solutions corresponding to simulated stress (salinity: 100 mM NaCl; drought: 10% PEG-6000) and exogenous exposure to phytohormones (100 μM ABA; 100 μM MeJA). The control plants remained in the water. Cold stress was simulated by placing the plants in a refrigerator (4°C, without light); the controls were kept in the dark at 22°C. After 6 h and 24 h of stress/hormone exposure, roots and sprouts were collected from three randomly selected experimental and control plants and stored at –80°C.

In the experiment without light, the plants were covered with a light-tight box (experiment) (10:00); the controls were under illumination at 190 μM/(m^2·s) (ECCF, day/night – 16/8 h). After 6 h (16:00) and 24 h (at 10:00 the next day), the roots and leaves were collected from the experiment and the control (two biological replicates for each point) and stored at –80°C.

The collected samples were used for RNA/cDNA extraction and qPCR as described above.

RESULTS

Identification and structural characterization of garlic *AsCHS* genes

Chalcone synthases belong to the type III polyketide synthase family, consist of two conserved domains, Chal_sti_synt_N (PF00195.16) and Chal_sti_synt_C (PF02797.12), and catalyze, as a homodimer, the addition of three malonyl-CoA molecules to 4-coumaroyl-CoA to form chalcone [3, 14]. Each component of the dimer comprises an active site that catalyzes one or more condensation reactions [14]. The CHS catalytic site contains four highly conserved amino acid residues (Cys164, His303, Asn336, and Phe215 in CHS1 *Glycin max*) [7, 15], where Cys164 acts also as the binding site for the 4-coumaroyl-CoA substrate [14]. Gly259 and Ser345 are involved in the binding of 4-coumaroyl-CoA, and a 17 aa consensus sequence is involved in the binding of the malonyl-CoA substrate [10]. In addition, a 17 aa signature sequence in the Chal_sti_synt_C domain has been proposed for chalcone synthases [10].

In the garlic *A. sativum* cv. Ershuizao genome [31], we identified eight chalcone synthase genes, *AsCHS1–8* (1,182–2,010 bp), that contained from one to three exons and were located on chromosomes Chr1 (*AsCHS1*), Chr4 (*AsCHS2, 3*), Chr5 (*AsCHS4, 5*), and Chr6 (*AsCHS6–8*) (*Table 2*) (*Fig. 1A*).

AsCHS1–8 proteins differed slightly in size (375–397 aa). According to functional predictions in Gene Ontology (GO) terms, all AsCHS1–8 exhibit acyltransferase activity (GO:0016747) and are involved in polyketide (GO:0030639) and flavonoid (GO:0009813) biosynthesis. In this case, AsCHS2 and 7 proteins had 75% similarity and differed significantly from AsCHS1, 3–6, and 8 chalcone synthases (56–61% identity).

The structural analysis of AsCHS1–8 amino acid sequences revealed the position of the chalcone synthase domains Chal_sti_synt_N (PF00195.16) and Chal_sti_synt_C (PF02797.12) (*Fig. 1*). Conserved residues (Cys167, Phe218 (Chal_sti_synt_N); His309, Asn342 (Chal_sti_synt_C)), characteristic of the enzyme's active site [15], were found in the domains, with the exception of AsCHS3 (Phe218→Cys). The 4-coumaroyl-CoA binding sites, Cys167 [14] and Gly259 and Ser345 [10], were present in all AsCHS1–8, with the exception of the Gly259→Lys mutation in AsCHS6. In the Chal_sti_synt_C domain, all

Table 2. Characterization of the chalcone synthase genes in the genome of garlic *A. sativum* cv. Ershuizao

Gene[1]	Gene/transcript ID[2]	Genome localization[2]	Gene, bp	Exon/ intron number	cDNA, bp	Protein, aa	MW, kDa	pI	GRAVY
AsCHS1	Asa1G03363.1/ Asa2G01293.1	chr1: 913472045..913473226	1,182	1/0	1,182	393	43.26	6.22	–0.139
AsCHS2	Asa4G02924.1/ Asa4G00890.1	chr4: 781174614..781176037	1,424	3/2	1,128	375	41.06	6.96	–0.084
AsCHS3	Asa4G06151.1/ Asa4G03387.1	chr4: 1682101557..1682102738	1,182	1/0	1,182	393	43.15	6.48	–0.121
AsCHS4	Asa5G04529.1/ Asa5G01644.1	chr5: 1227135621..1227136805	1,185	1/0	1,185	394	43.46	6.1	–0.179
AsCHS5	Asa5G04530.1/ Asa5G01645.1	chr5: 1227252338..1227253522	1,185	1/0	1,185	394	43.43	6.1	–0.177
AsCHS6	Asa6G02586.1/ Asa6G05452.1	chr6: 656216362..656217943	1,582	3/2	1,173	390	43.32	5.75	–0.126
AsCHS7	Asa6G03080.1/ Asa1G04064.1	chr6: 787943706..787944950	1,245	2/1	1,155	384	42.27	5.57	–0.155
AsCHS8	Asa6G03715.1/ Asa6G04348.1	chr6: 973362901..973364911	2,010	2/1	1,194	397	43.78	6.48	–0.186

[1]The numbers in the gene names are assigned in the order of their chromosomal location.
[2]Derived from garlic genome and transcriptome sequencing data [31].

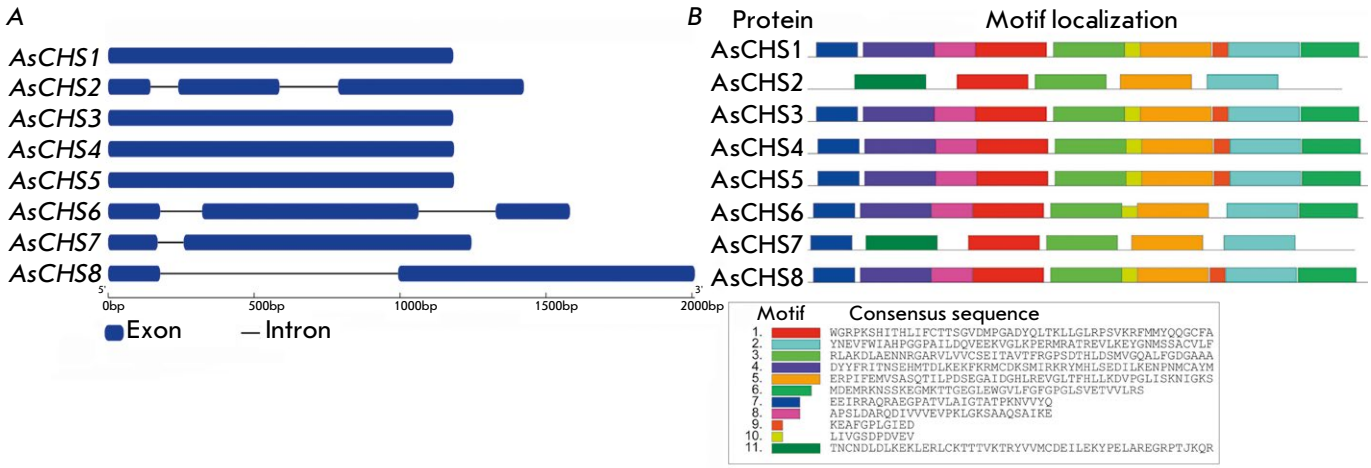


Fig. 1. The exon–intron structures of *AsCHS1–8* genes (*A*) and the composition and distribution of conserved motifs in AsCHS1–8 protein sequences (*B*)

AsCHS1–8 contained the malonyl-CoA binding consensus and the chalcone synthase signature sequence (*Fig. 2*).

The AsCHS1–8 sequences were characterized in terms of their conserved motif/consensus profiles (*Fig. 1B*). Most of the chalcone synthases (AsCHS1, 3–5, 8) contained 10 of the 11 identified motifs; AsCHS6 differed only in the lack of motif 9. The exceptions were AsCHS2 and AsCHS7 (five and six motifs, respectively, instead of 10 or 11); however, the sequences of these proteins specifically contained a motif 11 that corresponded to an altered (in comparison with other proteins) beginning of the Chal_sti_synt_N domain; AsCHS2 lacked consensus 7 due to a deletion at the beginning of the Chal_sti_synt_N domain. Motifs 6 and 8–10 were lost in AsCHS2 and AsCHS7, because the conservation of these regions was <50% in comparison with that in other chalcone synthases (*Figs. 1, 2*).

To investigate the phylogeny of garlic chalcone synthases AsCHS1–8, we used the AlliumDB and NCBI databases to identify the sequences of these enzymes in the species most closely related to *A. sativum*: onion *A. cepa* (6 CHSs), Welsh onion *A. fistulosum* (5), and asparagus *As. officinalis* (6), as well as in distant species: pepper *C. annuum* (9), tomato *S. lycopersicum* (7), and *A. thaliana* (4).

In the constructed phylogenetic tree (*Fig. 3*), AsCHS1–8 proteins were grouped with representatives of other monocot species (*A. cepa*, *A. fistulosum*, *As. officinalis*). Orthologs of AsCHS6–8 were found in all three species, orthologs of AsCHS2 were discov-

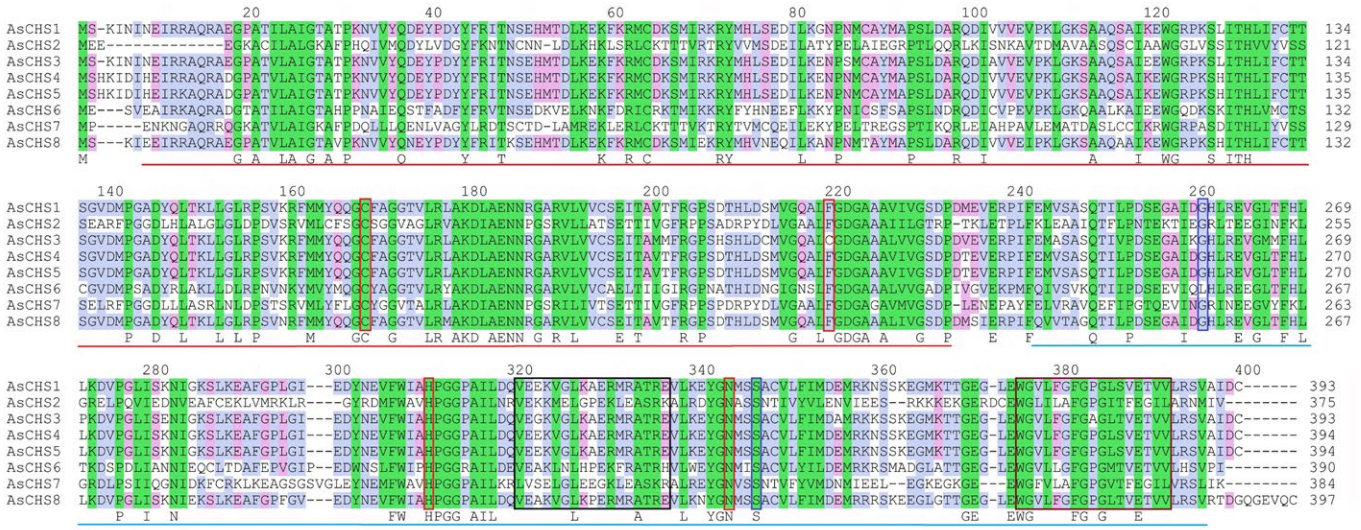


Fig. 2. Alignment of AsCHS1–8 amino acid sequences. The Chal_sti_synt_N and Chal_sti_synt_C domains are underlined by red and blue lines, respectively. Four residues (Cys167, Phe218, His309, Asn342) of the enzyme's active site are framed in red (according to [15]). Residues Ser345 and Gly259, involved in the binding of the 4-coumaroyl-CoA substrate [10], are framed in blue. The black and brown frames highlight the chalcone synthase malonyl-CoA binding consensus and signature sequences, respectively [10]. The background color indicates a high degree of amino acid conservation in AsCHS1–8 proteins (green – 100%, blue – 80%, and pink – 60%)

ered only in *A. cepa*, whereas orthologs of AsCHS1, 3–5 were grouped separately from representatives of both monocots and dicots (*Fig. 3*).

Only orthologs of garlic chalcone synthases AsCHS2 and 7 were found in dicot (*C. annuum*, *S. lycopersicum*, *A. thaliana*) genomes (*Fig. 3*).

Identification of *AsCHS* gene expression patterns in garlic plants

Chalcone synthase gene expression patterns were determined using transcriptomic data for individual organs of *A. sativum* cv. Ershuizao [31], including eight stages of bulb development (*Fig. 4*).

The *AsCHS2* and *AsCHS7* genes were found not to be expressed, except for trace numbers of transcripts in flowers (both genes), roots, and bulbs at certain developmental stages (*AsCHS7*). Expression of the remaining six genes was extremely insignificant in the roots, leaves, pseudostems, flowers, and during bulb development (*AsCHS1*, *3–6*), as well as in buds (except *AsCHS5*) and sprouts (except *AsCHS1*). Among *AsCHS1–7* genes, despite their low expression levels, *AsCHS3* may be detected (FPKM is significantly higher than that of the other five genes, but <10) (*Fig. 4*).

Significant transcript levels (FPKM >10) were found only for the *AsCHS8* gene. The *AsCHS8* gene is expressed in all analyzed organs, with the highest FPKM values being present in the pseudostem, leaves, flowers, and sprouts; in the bulb, expression is minimal throughout all eight developmental stages; in the roots, FPKM is ~9-fold lower than that in the leaves (*Fig. 4*).

Using qPCR, we determined the expression profiles of *AsCHS1–8* genes in the roots, basal plate (modified stem), bulb, pseudostem, and leaves of garlic (cv. Sarmat). Expression of two of the eight genes, *AsCHS3* and *8*, was detected. *AsCHS3* transcripts were present in all analyzed organs (maximum in the bulb, leaves, and pseudostem), whereas *AsCHS8* was only expressed in the roots, leaves, and pseudostem (maximum in the pseudostem and leaves). In the roots, both genes were expressed at a trace level but expression of *AsCHS3* was 26-fold lower than that of *AsCHS8*. In the pseudostem and leaves, the *AsCHS8* transcript levels were ~41-fold higher than those of *AsCHS3* (*Fig. 5A*).

Dynamics of *CHS* gene expression in garlic plants in response to stressors and phytohormones

To evaluate the possible involvement of chalcone synthases AsCHS1–8 in the stress response of garlic cv. Sarmat, we conducted a series of experiments to simulate the effects of salinity, drought, low positive temperatures, as well as exogenous treatment with ab-

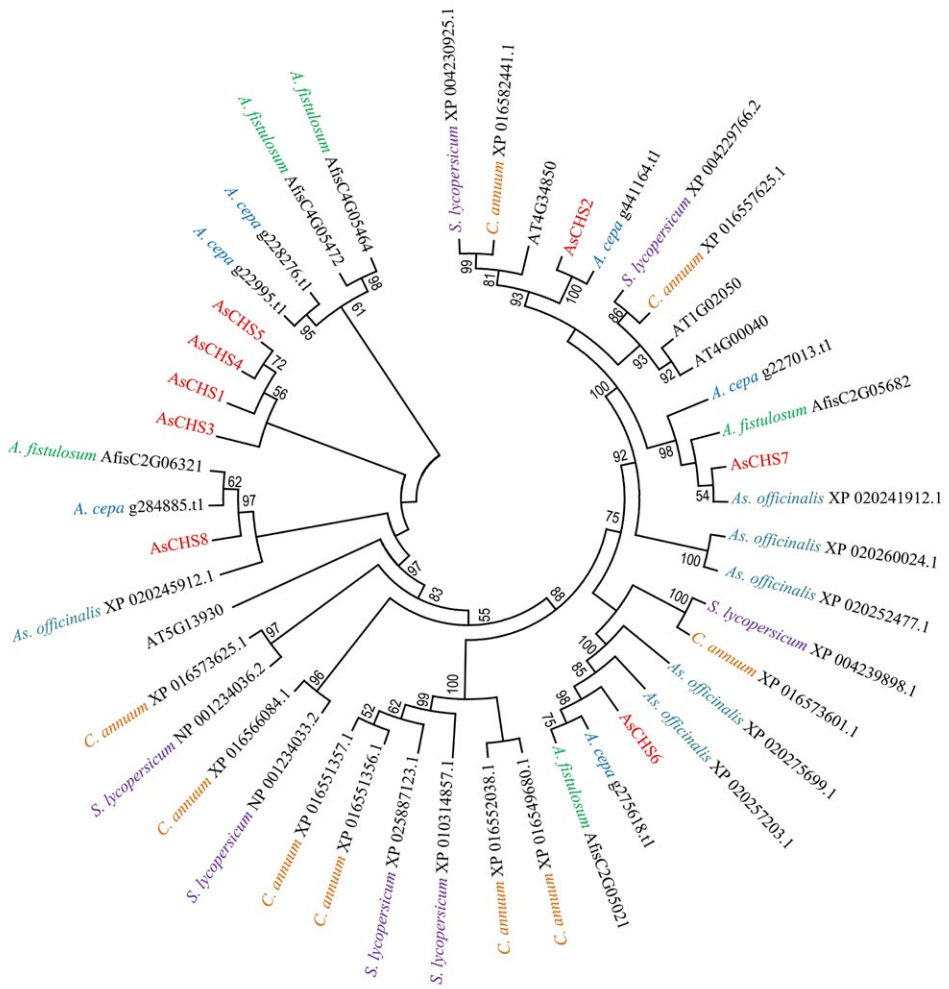


Fig. 3. The phylogenetic tree based on the amino acid sequences of chalcone synthases from *A. sativum* (As; red font), *A. cepa* (blue), *A. fistulosum* (green), *As. officinalis* (light blue), *A. thaliana* (AT; black), *C. annuum* (orange), and *S. lycopersicum* (purple). Significant bootstrap values (>50%) are indicated below the branches; the branch length corresponds to the number of mutations during evolution

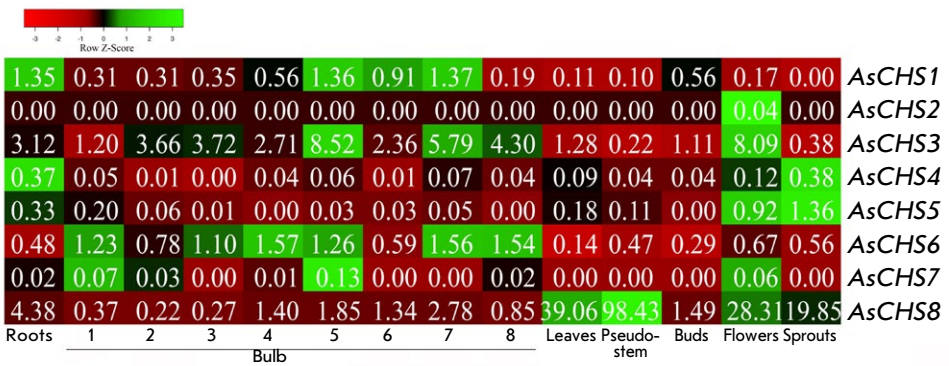


Fig. 4. The heat map of *AsCHS1–8* gene expression in different organs of *A. sativum* cv. Ershuizao, constructed using transcriptomic data [31]. The numbers in rectangles indicate the average FPKM value derived from three biological replicates. The development stages (1–8) of bulbs are shown; the age of the bulbs is 192, 197, 202, 207, 212, 217, 222, and 227 days, respectively [31]

scisic acid and methyl jasmonate on the plants and analyzed the expression of *AsCHS1–8* genes in the roots and leaves over time.

Only the *AsCHS8* gene was shown to be significantly expressed, both in the control and in the experiment (*Fig. 5B*). Trace amounts of *AsCHS2–4* transcripts were detected in the leaves in response to cold stress (figures are not shown, and the effect is not discussed due to the insignificance of gene induction).

The expression pattern of the *AsCHS8* gene in response to stress depends on both the type of stressor and the plant organ (roots or leaves).

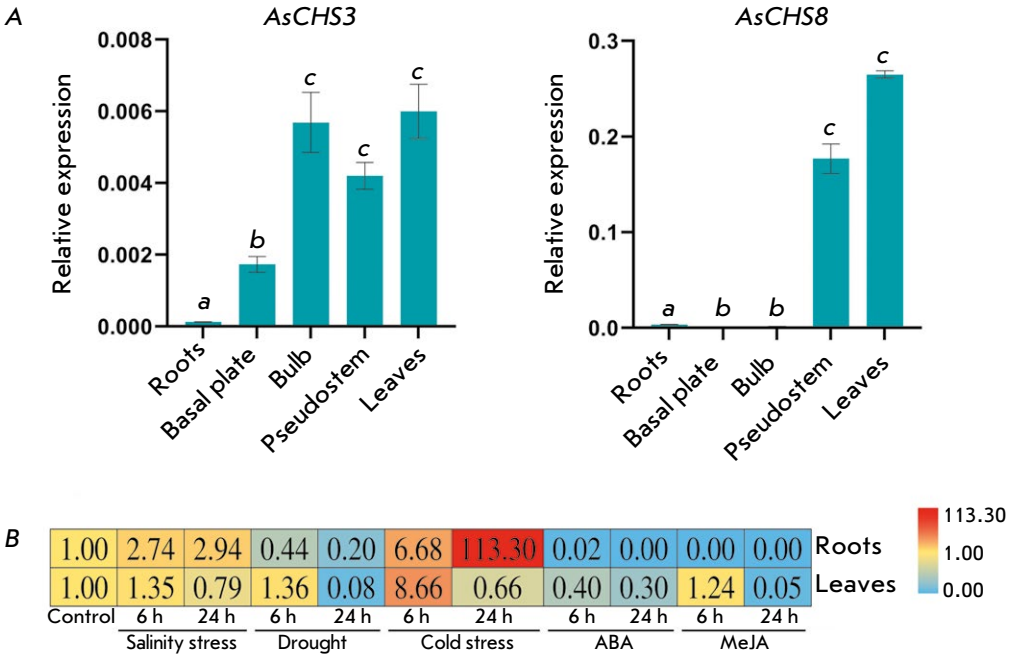


Fig. 5. (*A*) The expression patterns (qPCR) of the *AsCHS3* and *8* genes in different organs of adult garlic plants (cv. Sarmat) ($^{a-c}p < 0.05$ – significant differences in expression levels in different organs). Expression of the *AsCHS1*, *2*, and *4–7* genes was not detected (not shown in the figure). (*B*) Changes in *AsCHS8* expression levels in the roots and leaves of garlic plants in response to abiotic stresses (salinity, drought, cold) and exogenous phytohormones (ABA, MeJA)

Excess salt mainly stimulates *AsCHS8* expression in both roots and leaves. In the roots, gene expression increases 2.7-fold (6 h) and 2.9-fold (24 h) compared with that in the controls; in the leaves, it increases 1.3-fold (6 h), but decreases 1.2-fold by the end of exposure (24 h) (*Fig. 5B*).

Under drought conditions, *AsCHS8* expression in the roots steadily decreases (6 and 24 h), whereas it is initially activated 1.3-fold (6 h) and then sharply decreases to almost zero (24 h) in the leaves (*Fig. 5B*).

Cold stress stimulates *AsCHS8* expression in the roots (6.6-fold) and leaves (8.6-fold) at the beginning of exposure (6 h). At the end of exposure (24 h), gene transcript levels in the roots are increased 113.3-fold, whereas they are down 1.5-fold in the leaves (*Fig. 5B*).

Therefore, regarding changes in gene expression, all three types of abiotic stressors at comparable levels affect *AsCHS8* expression in the leaves, whereas the effect on roots is specific for each stressor.

Exogenous treatment of garlic plants with abscisic acid and methyl jasmonate completely suppresses *AsCHS8* expression in the roots. In the leaves, ABA significantly reduces gene expression within 24 h, whereas MeJA initially increases transcript levels 1.2-fold (6 h) and then suppresses them down to trace amounts (24 h) (*Fig. 5B*).

In addition, we analyzed the dependence of *AsCHS8* expression on illumination of the roots and leaves of plants placed in standard light (control) and dark (experiment) conditions (*Fig. 6*). We found that the *AsCHS8* gene was expressed in the roots of the

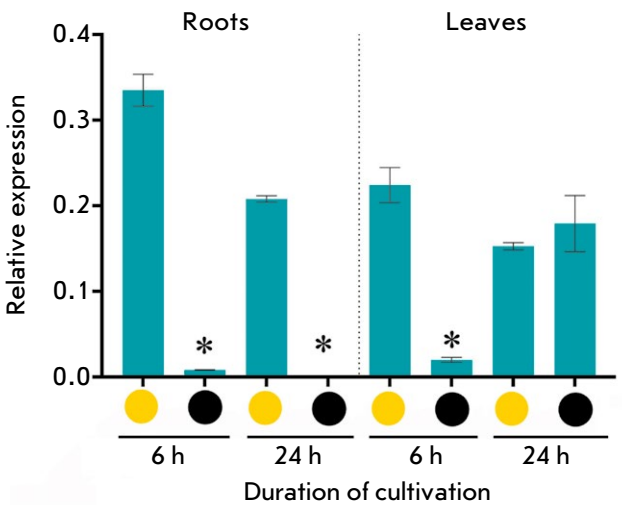


Fig. 6. Expression of the *AsCHS8* gene in the roots and leaves of garlic plants after 6 and 24 h of cultivation under illumination (yellow circle) and in the dark (black circle). (*$p < 0.05$ – significant differences in expression levels, darkness *vs.* light)

control (illuminated) plants, whereas trace amounts of transcripts could be detected in the darkness only after 6 h of exposure. In the leaves, *AsCHS8* was expressed in both the control and the experiment: after 6 h, gene transcript levels under dark conditions had dropped 11.2-fold compared to those under illumination; after 24 h, *AsCHS8* transcript levels were similar in the control and experiment (*Fig. 6*).

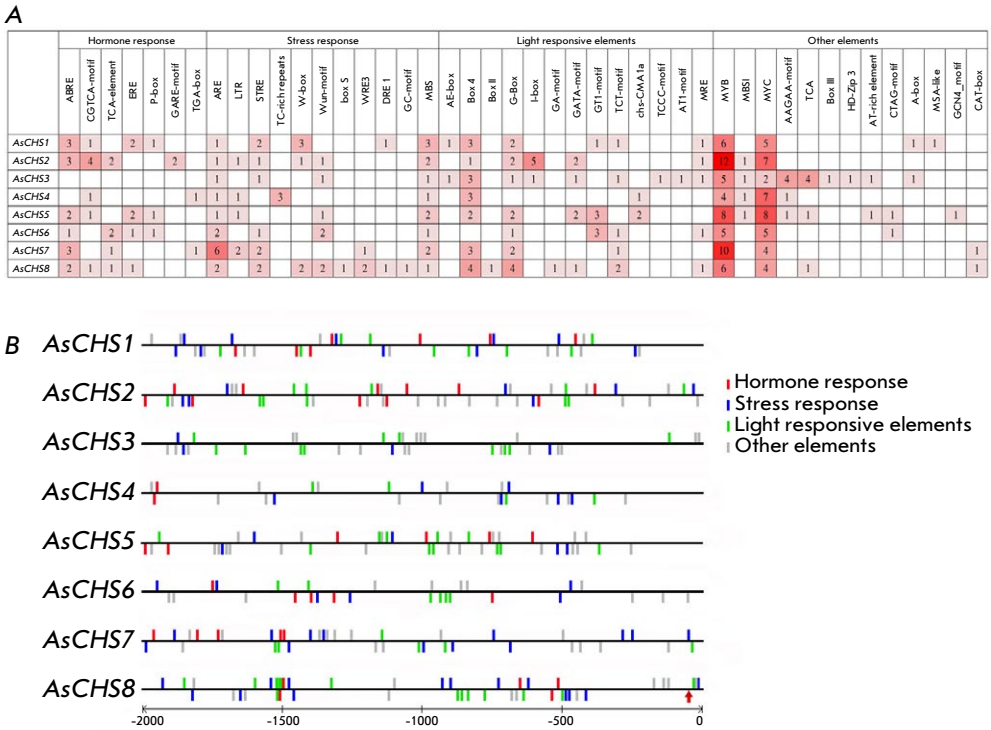


Fig. 7. The content and composition of *cis*-regulatory elements in the promoters (2 kb upstream of the start codon, including the putative 5′-untranslated region (5′-UTR)) of *AsCHS1–8* genes, (*A*) and their distribution along the promoter sequence (*B*). The red arrow indicates the putative *AsCHS8* transcription start site (based on transcriptomic data)

Identification of *cis*-regulatory elements in *AsCHS1–8* gene promoters

In order to interpret the expression patterns of *AsCHS1–8* genes, their promoters (-2,000 bp upstream of the start codon) were characterized by the profile of *cis*-regulatory elements (*Fig. 7*). We found 44 elements, and they were divided into four groups: phytohormone (7) and stress (11) response elements, as well as light-sensitive (13) and other (13) elements. The latter include binding sites for proteins and transcription factors, the development-associated element, and potential regulatory motifs with an unknown function.

The promoters of most genes (except *AsCHS3*) were shown to contain phytohormone-sensitive elements, with the abscisic acid and methyl jasmonate responsive sites predominating among them. The *AsCHS1*, *2*, and *7* genes contain the largest number of ABA responsive elements (3 'ABRE'); *AsCHS2* contains the largest number of MeJA responsive elements (4 'CGTCA'). Auxin- ('TGA'), gibberellin- ('P-box', 'GARE'), salicylic acid- ('TCA'), and ethylene-associated ('ERE') elements are present in the promoters of individual genes in one or two copies (*Fig. 7*).

In the promoters of all *AsCHS1–8* genes, we identified elements associated with the response to stress in general ('MBS', 'W-box', 'TC-rich repeats'), anaerobic conditions ('ARE', especially *AsCHS7* with 6 sites), phytopathogens ('Wun', 'WRE3', 'box S'), cold ('LTR'), drought ('DRE1'), osmotic stress, heat, and nutrients deficiency ('STRE'). The largest number of elements (14) was found in *AsCHS8*. Given the stress factors used in this study, we note that 'LTR' elements (cold response) were found in *AsCHS2*, *4*, *5*, and *7*; 'DRE1' and/or 'STRE' (osmotic stress response), in all genes, except *AsCHS4* and *5* (*Fig. 7*).

Our analysis revealed that the promoter regions of *AsCHS1–8* contain from 4 (*AsCHS4*) to 14 (*AsCHS8*) light-sensitive elements (*Fig. 7*).

In all the genes, we found binding sites for the stress-associated transcription factors of the MYB ('MRE', 'MYB', 'MBS1') and MYC ('MYC') families: 5–14 and 2–8 elements, respectively. The MYB-binding sites are the most enriched in the promoters of the *AsCHS2* (14), *7* (10), and *4* (9) genes; *AsCHS2–5* promoters contain the 'MBS1' associated with flavonoid biosynthesis regulation (*Fig. 7*).

DISCUSSION

The plant's defense response is associated with the accumulation of metabolites that possess antioxidant properties; in particular flavonoids [5, 6]. Garlic *A. sativum*, which lost its fertility during evolution and domestication, has seriously altered its genetic regulation of stress adaptation [34]. The flavonoid pathway in garlic has not been characterized. Therefore, the aim of this study was to identify and structurally and functionally characterize the *A. sativum* genes encod-

ing chalcone synthases that catalyze the first stage of the flavonoid biosynthesis pathway [12].

The analysis of the genome and transcriptomes of *A. sativum* cv. Ershuizao revealed eight *AsCHS1–8* chalcone synthase genes (*Table 2*). The number of genes from this family in the garlic genome was different from that in other monocots, such as wheat *T. aestivum* (49 or 87 genes) or maize *Z. mays* (17). However, the sizes of this family in *A. sativum* and in one of the species most closely related to the genus *Allium*, *As. officinalis* (six genes), were found to be comparable [18, 21]. Since the high degree of phenotypic variability of garlic today is believed to result from the cross-breeding of the fertile wild ancestors at the center of the origin of the species [25, 34], one could suggest that the *AsCHS* family had appeared in the garlic genome even before the species lost its ability to reproduce sexually.

Given the results of the structural and phylogenetic analysis (*Table 2*, *Figs. 1* and *3*), it is fair to suggest that the highly homologous proteins AsCHS1, 3–5 are functionally redundant and can function in partial overlap in different plant tissues/organs, this being controlled by the specificity of the gene promoters. The corresponding genes differ significantly in their set of *cis*-regulatory elements in the promoter region (*Fig. 7*) and the organ-specific expression pattern (*Fig. 4*) derived from the transcriptomic data of *A. sativum* cv. Ershuizao [31]. The involvement of genes in the flavonoid biosynthesis may be limited to the individual, highly specialized cells/tissues/organs/developmental stages of the garlic plant. The identified mutations in the essential amino acid residues in AsCHS3 and 6 (*Fig. 2*), which are required for formation of the substrate binding site, may also be an expression of possible differences in the enzymatic activity of these proteins [10, 15].

In general, as regards the expression of all the analyzed *AsCHS* genes (*Fig. 4*), significant expression of only *AsCHS8* (FPKM >10) and, to a lesser extent, *AsCHS3* (*Fig. 4*) deserves note. This is further confirmed by qPCR results indicating that only *AsCHS3* and *8* genes are expressed, with *AsCHS8* transcripts significantly predominating in the roots, pseudostem, and the leaves of garlic cv. Sarmat (*Fig. 5A*).

The lack or low expression of the other *AsCHS1*, *2*, and *4–7* genes does not yet constitute evidence of their dysfunction. All those genes remain structurally intact, including the profile of *cis*-regulatory elements in the promoter region (*Table 2*, *Fig. 7*); they can be highly specialized, participating in the flavonoid pathway in specific cells/tissues/organs at certain stages of plant development. For example, a number of wheat chalcone synthase genes are expressed exclusively in other cells during pollen exine development [21]. We, in addition, analyzed the expression of the *AsCHS1–8* genes in response to the main abiotic stressors (salinity, drought, cold) and exogenous treatment with phytohormones (abscisic acid and methyl jasmonate) that mediate the stress response signaling pathways in plants [35]. We found that only the *AsCHS8* gene expression was significantly altered in the response to all the stressors used (*Fig. 5B*).

The demonstrated stimulating effect of cold on *AsCHS8* gene activity (*Fig. 5B*) is consistent with data in similar studies conducted, for example, on *Coelogyne ovalis* [36] or *Oryza sativa* [37] plants. The increase in *AsCHS8* gene expression in response to salinity (*Fig. 5B*) is consistent with data on the response of rice plants [37] and the positive association between chalcone synthase gene expression and salt tolerance in *Eupatorium adenophorum* plants [38].

In contrast to the effects of salt and cold, the response to another osmotic stress factor, drought, is accompanied by a decrease in *AsCHS8* gene expression (*Fig. 5B*). On the one hand, this is consistent with data on *Camellia sinensis* plants known to decrease their chalcone synthase content in response to drought [39]. On the other hand, the effect of *AsCHS8* is contrary to the response of three chalcone synthase genes from *Silybum marianum,* whose expression has been shown to increase in response to drought [40].

It is worth noting that treatment of garlic plants with abscisic acid and methyl jasmonate suppresses the expression of the *AsCHS8* gene in both the roots and leaves (*Fig. 5B*). On the contrary, in *Vitis sp.* species, these treatments stimulate the expression of chalcone synthase genes [41, 42]; in the case of MeJA, this is associated with the activation of the biosynthesis of antimicrobial phytoalexins by jasmonates to protect against pathogens [41]. Probably, the opposite effect is due to the fact that garlic plants are rich in biologically active organosulfur compounds that exhibit strong antioxidant and antimicrobial properties [43] and that the flavonoid synthesis triggered by the jasmonate signaling pathway in other plant species is not that material to the defense response. Furthermore, treatment of *Salvia miltiorrhiza* plants with MeJA has been shown to either stimulate (*SmCHS1–5*), suppress (*SmCHS6*), or have no effect (*SmCHS7*) on the expression of chalcone synthase genes [44].

Flavonoids are known to be involved in plant photoprotection [1, 2]. In this case, flavonoid biosynthesis is positively dependent on illumination [45], which is associated with light-sensitive *cis*-regulatory elements in the promoters of chalcone synthase genes [7, 46]. We also found a significant number of light-sensitive

sites in the promoters of all *AsCHS* genes, in particular *AsCHS8* (*Fig. 7*), which is consistent with the demonstrated suppression of *AsCHS8* expression in plants in the dark (*Fig. 6*). A similar expression pattern of chalcone synthase genes, under illumination or in the dark, is typical of other plant species, e.g., *Sinapis alba* [45].

Thus, only one gene (*AsCHS8*) in the entire *AsCHS* family is involved in the defense response in garlic leaves and roots, with the response dynamics of gene expression depending on the nature of the stressor and being often in opposite direction. These data may be indirect confirmation that, during evolution and domestication, garlic plants have undergone serious changes in their genetic regulation of adaptation to stress, changes that are different from those that took place in other plant species [34]. Therefore, further research in this direction is required.

CONCLUSION

We identified and characterized eight chalcone synthase genes (*AsCHS1–8*) in the garlic *A. sativum* cv. Ershuizao genome, compared their organ-specific expression patterns with those in the cultivar Sarmat, and analyzed gene expression in response to abiotic stressors (salinity, drought, cold), exogenous phytohormones (ABA, MeJA) (all genes), and illumination (*AsCHS8* only). Our findings suggest that only two genes out of the eight, *AsCHS3* and *8*, are able to control flavonoid synthesis in all the analyzed organs during garlic plant development, and that the main chalcone synthase activity is determined by *AsCHS8*, whose expression in individual organs is not only the most significant, but also the most sensitive to stress factors. The other six genes (*AsCHS1*, *2*, *4–7*) may be involved in flavonoid biosynthesis in highly specialized cells/tissues/organs or at certain stages of garlic plant development. The identification and characterization of garlic chalcone synthase genes *AsCHS1–8* may form the basis for further analysis of the mechanisms that regulate stress adaptation in *A. sativum* and other *Allium* species. ●

The authors declare no conflict of interest.

This study was supported by a grant from the Russian Science Foundation (No. 24-76-10005; gene characterization and analysis, expression analysis) and the Ministry of Science and Higher Education of the Russian Federation (transcriptome-based expression analysis).

This article does not contain any studies that used animals or humans as research subjects.

REFERENCES

1. Taylor L.P., Grotewold E. // Curr. Opin. Plant Biol. 2005. V. 8. P. 317–323. doi: 10.1016/j.pbi.2005.03.005.
2. Mouradov A., Spangenberg G. // Front. Plant Sci. 2014. V. 5. Article 620. doi: 10.3389/fpls.2014.00620.
3. Petrussa E., Braidot E., Zancani M., Peresson C., Bertolini A., Patui S., Vianello A. // Int. J. Mol. Sci. 2013. V. 14. P. 14950–14973. doi: 10.3390/ijms140714950.
4. Shen N., Wang T., Gan Q., Liu S., Wang L., Jin B. // Food Chem. 2022. V. 383. Article 132531. doi: 10.1016/j.foodchem.2022.132531.
5. Han Y.Y., Ming F., Wang W., Wang J.W., Ye M.M., Shen D.L. // Genetica. 2006. V. 128. P. 429–438. doi: 10.1007/s10709-006-7668-x.
6. Khlestkina E. // Cereal Res. Commun. 2013. V. 41. P. 185–198. doi: 10.1556/CRC.2013.0004.
7. Dao T.T., Linthorst H.J., Verpoorte R. // Phytochem. Rev. 2011. V. 10. № 3. P. 397–412. doi: 10.1007/s11101-011-9211-7.
8. Shih C.H., Chu H., Tang L.K., Sakamoto W., Maekawa M., Chu I.K., Wang M., Lo C. // Planta. 2008. V. 228. № 6. P. 1043–1054. doi: 10.1007/s00425-008-0806-1.
9. Liu W., Feng Y., Yu S., Fan Z., Li X., Li J., Yin H. // Int. J. Mol. Sci. 2021. V. 22. № 23. Article 12824. doi: 10.3390/ijms222312824.
10. Zhu L., Ding Y., Wang S., Wang Z., Dai L. // Genes. 2022. V. 13. Article 2145. doi: 10.3390/genes13112145.
11. Abe I., Morita H. // Nat. Prod. Rep. 2010. V. 27. № 6. P. 809–838. doi: 10.1039/b909988n.
12. Singh B., Kumar A., Malik A.K. // Electrophoresis. 2017. V. 38. P. 820–832. doi: 10.1002/elps.201600334.
13. Imaizumi R., Mameda R., Takeshita K., Kubo H., Sakai N., Nakata S., Takahashi S., Kataoka K., Yamamoto M., Nakayama T., et al. // Proteins. 2020. V. 89. № 1. P. 126–131. doi: 10.1002/prot.25988.
14. Jez J.M., Bowman M.E., Noel J.P. // Biochemistry. 2001. V. 40. № 49. P. 14829–14838. doi: 10.1021/bi015621z.
15. Noel J.P., Ferrer J.-L., Jez J.M., Bowman M.E., Dixon R.A. // Nat. Struct. Biol. 1999. V. 6. P. 775–784. doi: 10.1038/11553.
16. Wu X., Zhang S., Liu X., Shang J., Zhang A., Zhu Z., Zha D. // PLoS One. 2020. V. 15. № 4. Article e0226537. doi: 10.1371/journal.pone.0226537.
17. Jia Y.H., He F., Shen Z.L., Xie X.H., Lv S.J., Jiang B.X., Yang G.X., Yan Y.C., Wu Z.H., Wu Y.Y. // Gene. 2023. V. 857. Article 147176. doi: 10.1016/j.gene.2023.147176.
18. Yang L., Zhang S., Chu D., Wang X. // Front. Genet. 2024. V. 15. Article 1368358. doi: 10.3389/fgene.2024.1368358.
19. Han Y., Ding T., Su B., Jiang H. // Int. J. Mol. Sci. 2016. V. 17. № 2. Article 161. doi: 10.3390/ijms17020161.
20. Glagoleva A.Y., Ivanisenko N.V., Khlestkina E.K. // BMC Genet. 2019. V. 20. Article 30. doi: 10.1186/s12863-019-0727-y.
21. Liu Y., Bai J., Yuan S., Gao S., Liu Z., Li Y., Zhang F., Zhao C., Zhang L. // Gene. 2023. V. 888. Article 147740. doi: 10.1016/j.gene.2023.147740.
22. Deng X., Bashandy H., Ainasoja M., Kontturi J., Pie-

tiäinen M., Laitinen R.A.E., Albert V.A., Valkonen J.P.T., Elomaa P., Teeri T.H. // New Phytol. 2014. V. 201. № 4. P. 1469–1483. doi: 10.1111/nph.12610.
23. Des Marais D.L., Rausher M.D. // Nature 2008. V. 454. P. 762–765. doi: 10.1038/nature07092.
24. El-Saber Batiha G., Magdy Beshbishy A., Wasef L.G., Elewa Y.H.A., Al-Sagan A.A., Abd El-Hack M.E., Taha A.E., Abd-Elhakim Y.M., Prasad Devkota H. // Nutrients. 2020. V. 12. № 3. Article 872. doi: 10.3390/nu12030872.
25. Shemesh E., Scholten O., Rabinowitch H.D., Kamenetsky R. // Planta. 2008. V. 227. P. 1013–1024. doi: 10.1007/s00425-007-0675-z.
26. Kıraç H., Dalda Şekerci A., Coşkun Ö.F., Gülşen O. // Genet. Resour. Crop Evol. 2022. V. 69. P. 1833–1841. doi: 10.1007/s10722-022-01343-4.
27. Casals J., Rivera A., Campo S., Aymerich E., Isern H., Fenero D., Garriga A., Palou A., Monfort A., Howad W., et al. // Front. Plant Sci. 2023. V. 13. Article 1004069. doi: 10.3389/fpls.2022.1004069.
28. Buitrago S., Yang X., Wang L., Pan R., Zhang W. // Plant Mol. Biol. 2024. V. 115. № 1. Article 6. doi: 10.1007/s11103-024-01540-y.
29. Kim S., Yoo K.S., Pike L.M. // Euphytica. 2005. V. 142. P. 273–282. doi: 10.1007/s10681-005-2239-2.
30. Mollavali M., Perner H., Rohn S., Riehle P., Hanschen F.S., Schwarz D. // Mycorrhiza. 2018. V. 28. № 1. P. 59–70. doi: 10.1007/s00572-017-0799-3.
31. Sun X., Zhu S., Li N., Cheng Y., Zhao J., Qiao X., Lu L., Liu S., Wang Y., Liu C., et al. // Mol. Plant. 2020. V. 13. № 9. P. 1328–1339. doi: 10.1016/j.molp.2020.07.019.
32. Liu M., Wu Z., Jiang F. // Plant Cell Tiss. Organ Cult. 2015. V. 122. P. 435–444. doi: 10.1007/s11240-015-0780-9.
33. Schwinn K.E., Ngo H., Kenel F., Brummell D.A., Albert N.W., McCallum J.A., Pither-Joyce M., Crowhurst R.N., Eady C., Davies K.M. // Front. Plant. Sci. 2016. V. 7. Article 1865. doi: 10.3389/fpls.2016.01865.
34. Shemesh-Mayer E., Faigenboim A., Sherman A., Gao S., Zeng Z., Liu T., Kamenetsky-Goldstein R. // Int. J. Mol. Sci. 2023. V. 24. № 23. Article 16777. doi: 10.3390/ijms242316777.
35. Waadt R., Seller C.A., Hsu P.K., Takahashi Y., Munemasa S., Schroeder J.I. // Nat. Rev. Mol. Cell Biol. 2022. V. 23. № 10. P. 680–694. doi: 10.1038/s41580-022-00479-6.
36. Singh N., Kumaria S. // Gene. 2020. V. 762. Article 145104. doi: 10.1016/j.gene.2020.145104.
37. Wang J., Zhang C., Li Y. // Genes (Basel). 2022. V. 13. № 3. Article 410. doi: 10.3390/genes13030410.
38. Lijuan C., Huiming G., Yi L., Hongmei C. // Plant Cell Rep. 2015. V. 34. № 5. P. 885–894. doi: 10.1007/s00299-015-1751-7.
39. Gu H., Wang Y., Xie H., Qiu C., Zhang S., Xiao J., Li H., Chen L., Li X., Ding Z. // Sci. Rep. 2020. V. 10. № 1. Article 15504. doi: 10.1038/s41598-020-72596-1.
40. ElSayed A.I., El-Hamahmy M.A.M., Rafudeen M.S., Mohamed A.H., Omar A.A. // Plants (Basel). 2019. V. 8. № 12. Article 611. doi: 10.3390/plants8120611.
41. Nopo-Olazabal C., Condori J., Nopo-Olazabal L., Medina-Bolivar F. // Plant Physiol. Biochem. 2014. V. 74. P. 50–69. doi: 10.1016/j.plaphy.2013.10.035.
42. Yang M., Wang L., Belwal T., Zhang X., Lu H., Chen C., Li L. // Molecules. 2019. V. 25. № 1. P. 12. doi: 10.3390/molecules25010012.
43. Shang A., Cao S.Y., Xu X.Y., Gan R.Y., Tang G.Y., Corke H., Mavumengwana V., Li H.B. // Foods. 2019. V. 8. № 7. Article 246. doi: 10.3390/foods8070246.
44. Deng Y., Li C., Li H., Lu S. // Molecules. 2018. V. 23. № 6. Article 1467. doi: 10.3390/molecules23061467.
45. Ehmann B., Ocker B., Schafer E. // Planta. 1991. V. 183. P. 416–422. doi: 10.1007/BF00197741.
46. Hartmann U., Sagasser M., Mehrtens F., Stracke R., Weisshaar B. // Plant Mol. Biol. 2005. V. 57. № 2. P. 155–171. doi: 10.1007/s11103-004-6910-0.

The Emergence of a Novel Insertional Mutation in the *BCR::ABL*/p210 Oncogene in B-Cell Acute Lymphoblastic Leukemia (B-ALL) Correlates with the Development of Resistance to Several Tyrosine Kinase Inhibitors

K. V. Bogdanov*, E. S. Kudryavtseva, Y. N. Lobacheva, O. V. Merzlikina, Y. V. Mirolyubova, R. A. Vlasik, R. Sh. Badaev, E. G. Lomaia
Almazov National Medical Research Centre, St. Petersburg, 197341 Russia
*E-mail: kvbogdanov@yandex.ru
Received October 15, 2024; in final form, February 13, 2025
DOI: 10.32607/actanaturae.27539

ABSTRACT **A patient with an immunophenotype characteristic of B-cell acute lymphoblastic leukemia (B-ALL) was found to carry the chromosomal translocation t(9;22)(q34;q11), or Philadelphia (Ph) chromosome and less common variant of the chimeric oncogene *BCR::ABL*/p210. No additional mutations in the *BCR::ABL* gene, including point mutations, insertions, or deletions, were identified in the disease onset characterized by elevated blast cell (77.6%) and leukocyte (48×10^9/L) counts. Ph+ALL-2012m chemotherapy with imatinib (600 mg) and two consolidation phases resulted in complete hematologic remission and a profound molecular response. However, six months later, the patient had relapsed (blasts: 15%, *BCR::ABL*/p210: 105%). Three weeks after the initiation of dasatinib therapy (100 mg), the number of blasts had decreased to 4.8%, while the expression level of *BCR::ABL*/p210 had dropped to 11.8%. Sanger sequencing identified two mutations in the *BCR::ABL* oncogene; namely, the point mutation F317L and a new insertion of nine nucleotides previously not detected. In the latter case, the amino acid lysine at position 294 was replaced by four new amino acid residues: K294SPSQ. Therapy with bosutinib and inotuzumab led to the disappearance of one leukemia clone with the F317L mutation, but the presence of another clone carrying a nine-nucleotide insertion was observed. The switch to ponatinib+blinatumomab chemotherapy was effective, resulting in the disappearance of the insertion. Allogeneic hematopoietic stem cell transplantation (allo-HSCT) from an available HLA-matched unrelated donor resulted in complete clinical and hematologic remission, including a complete molecular response. Six months after allo-HSCT, minimal residual disease monitoring showed maintenance of complete remission.**
KEYWORDS **B-ALL, *BCR::ABL*/p210, insertion, K294SPSQ, resistance to tyrosine kinase inhibitors.**
ABBREVIATIONS **B-ALL – B-cell acute lymphoblastic leukemia; CML – chronic myeloid leukemia; TKI – tyrosine kinase inhibitor; IM – imatinib; allo-HSCT – allogeneic hematopoietic stem cell transplantation.**

INTRODUCTION

Acute B-lymphoblastic leukemia (B-ALL) is a clonal proliferative disease of the blood system caused by genetic abnormalities arising in B-cell precursor cells. The disease is more common in children and less so in adults. Among adult B-ALL patients, some mutations are rare (< 3%), namely: t(v;11q23)/MLL or KMT2A, including t(4;11)(q21;q23)/KMT2A-AF4, t(1;19)(q23;p13)/E2A-PBX1 (TCF3-PBX1), t(5;14)(q31;q32)/IL3-IGH, t(12;21)(p13;q22)/TEL-AML1 (ETV6-RUNX1), and C-MYC gene rearrangement mutations: t(8;14)(q24;q32), t(8;22)(q24;q11), or t(2;8)(p22;q23). In contrast, the Philadelphia chromosome (Ph chromosome) is a frequent mutation (up to 30%) among adult B-ALL patients. It results from reciprocal translocation t(9;22)(q34;q11), whose product is the chimeric *BCR::ABL* oncogene. Moreover, a variant of Ph-like B-ALL is usually identified among adult patients (⩾ 20%). It is characterized by a molecular gene expression profile typical of Ph-positive B-ALL but different by the absence of the chromosomal translocation t(9;22)(q34;q11) and the predominance of a

high frequency of deletions in the *IKZF1* gene [1]. Most of these mutations correlate with a poor prognosis, except for some chromosomal translocations: in particular t(1;19)(q23;p13)/E2A-PBX1 and t(12;21)(p13;q22)/TEL-AML1. These chromosomal mutations are associated with an intermediate and favorable prognosis, respectively. In addition to correlation with poor prognosis in B-ALL patients, the chromosomal translocation t(9;22)(q34;q11) leads to an accumulation of the BCR::ABL oncoprotein, with constitutive tyrosine kinase activity. The increased ability of BCR::ABL to phosphorylate target proteins causes a transformation of hematopoietic stem cells, resulting in the alteration of multiple signaling pathways that enhance their survival and proliferation [2–4]. Different fusion products can be detected depending on the localization of the breakpoint in the *BCR* gene or alternative splicing of *BCR-ABL* mRNA. The most common ones are the *BCR::ABL* isoforms: e1a2 (p190), e13a2, and e14a2 (both p210). Meanwhile, most B-ALL patients (77%) tend to express *BCR::ABL*/p190 while a smaller proportion of patients (20%) express *BCR::ABL*/p210, and the remainder (3%) co-express *BCR::ABL*/p190 and *BCR::ABL*/p210 [5]. Despite the generally unfavorable prognosis for Ph-positive B-ALL, the prognosis is even worse for *BCR::ABL*/p210 mutation carriers compared to that for *BCR::ABL*/p190 mutation carriers. The advent of imatinib (IM), a tyrosine kinase inhibitor (TKI) suppressing the BCR::ABL tyrosine kinase activity, has significantly improved the hematological, cytogenetic, and molecular genetic characteristics of Ph-positive patients [6]. However, the majority of Ph+ B-ALL patients often develop resistance to IM via both the BCR::ABL-dependent and BCR-ABL-independent mechanisms [7–12]. To overcome this resistance, second- (nilotinib, dasatinib, bosutinib), third- (ponatinib), and fourth-generation (asciminib) TKIs have been developed and introduced into clinical practice [13, 14]. The BCR::ABL-dependent factors contributing to resistance to TKIs include the mutations arising in the BCR::ABL gene that encodes the tyrosine kinase domain, including point mutations, insertions and deletions. The mentioned abnormalities can be detected both at the disease onset and during treatment. These mutations are rare (⩽ 12%) at the disease onset, but their detection may increase during TKI treatment and contribute to the emergence of resistance. More than 100 point mutations in the *BCR::ABL* oncogene are currently known and have been previously described for both B-ALL and chronic myeloid leukemia (CML) [15–17]. These mutations affect different regions of the BCR::ABL kinase domain. Among those (1) the phosphate-binding P-loop (P-loop); (2) the C-helix site responsible for allosteric regulation; (3) the ATP/IM binding site; (4) the catalytic site (SH2 contact, SH3 contact, C-loop); and (5) the activation loop (A-loop) (*Fig. 1*).

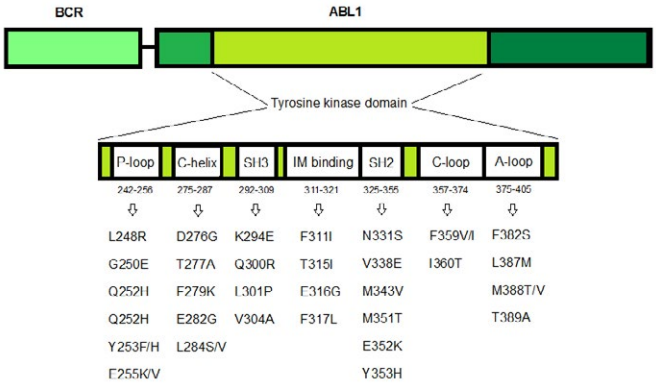


Fig. 1. BCR::ABL tyrosine kinase domain mutations identified in patients with Ph-positive leukemia

Mutations are most commonly found in two regions of the kinase domain, the P-loop and ATP/IM binding site. The most common of those is the T315I (C>T) point mutation. It substitutes threonine for isoleucine, causing resistance to four different TKIs namely imatinib, dasatinib, nilotinib, and bosutinib [18]. Among the insertional mutations in the BCR::ABL kinase domain in B-ALL patients, insertions of two to twelve amino acid residues, usually localized between positions I293 and K294, as well as K294 and H295, are more frequently detected. In both cases, the structure of the SH3 contact site, which constitutes the tyrosine kinase domain of the BCR::ABL oncoprotein, is disrupted. It leads to the development of resistance to imatinib [19, 20]. Finally, among the deletions of the *BCR::ABL* oncogene in B-ALL patients, a Δ184–274 mutation has been described, as being associated with the loss of 90 amino acid residues. In particular, this disruption affects the P-loop region of the BCR::ABL tyrosine kinase domain, which correlates with resistance to TKIs, including ponatinib. [21]. The present study has revealed an increased expression of the *BCR::ABL*/p210 oncogene and two mutations in the BCR::ABL kinase domain in a patient with B-ALL (Ph+) after dasatinib therapy. The first one, the F317L point mutation, is well-known; and the second mutation, a new insertion of nine nucleotides, has not been described previously. This insertion results in the substitution of a lysine at position K294 for four amino acid residues of SPSQ, which is part of the SH3-contact site of the tyrosine kinase domain of BCR::ABL.

Table 1. Main characteristics of the patient at the onset of B-ALL

Male	Peripheral blood						Bone marrow (blasts, %)
	hemoglobin, g/L	WBC, $\times 10^9$/L	platelets, $\times 10^9$/L	lymphocytes, %	monocytes, %	blasts, %	
Patient	101	48	30	15	3	67	77.6
Healthy person	130–160	4–9	150–400	19–37	3–11	0	0.1–1.1

Table 2. Specific oligonucleotides for PCR and direct sequencing of *BCR::ABL*

Primer	First round of PCR	Second round of PCR	Sequencing
Forward (5'-3')	ACTCGTGTGTGAAACTCCAGACT	AGGACGAGTATGCGCTGAAG	AGGACGAGTATGCGCTGAAG
Reverse (5'-3')	CGAGGTTTTGTGCAGTGAGC	CGAGGTTTTGTGCAGTGAGC	CGAGGTTTTGTGCAGTGAGC

EXPERIMENTAL

Patient and samples

A 42-year-old man was admitted to the Almazov National Medical Research Centre with a complain of pain in the knee joint and fever up to 38°C. According to the clinical analysis (*Table 1*), the patient had a high white blood cell count (48 × 10^9/L), as well as an increased blast count in the peripheral blood (67%) and bone marrow (77.6%). Immunophenotyping revealed a population of blast cells with the B-lineage phenotype: CD34+CD19+cytCD79a+CD10+CD38+sCD22+cytIgM-HLADR-CD13+MPO-CD33-CD117-. Hence, the patient was diagnosed with B-cell acute lymphoblastic leukemia (variant B II) with co-expression of the CD13+ myeloid marker. In addition, no involvement of the central nervous system was detected, since all the lumbar punctures showed the absence of leukemia cells in the cerebrospinal fluid.

Cytogenetic analysis

Preparation of chromosome spreads and subsequent chromosome differential staining were performed according to the assay described previously [22]. The karyotype pathology was interpreted by analyzing 20 mitoses by standard karyotyping and/or 200 interphase nuclei after fluorescence *in situ* hybridization (FISH).

RNA isolation, reverse transcription (RT), and quantitative real-time PCR (qPCR)

After the isolation of total RNA from 2.5 mL of peripheral blood and elution in 30 μL of RNase-free buffer, reverse transcription was performed using a standard set of reagents according to the manufacturer's protocol (AmpliSense, Russia). Qualitative determination of the *BCR::ABL* fusion transcript variant was performed using microchip PCR and a 5× qPCRmix-HS reagent kit (Eurogen, Russia) as described previously [23]. Quantification of the *BCR::ABL*/p210 oncogene was performed using qPCR and the Leucosis Quantum M-bcr-FRT PCR kit (AmpliSens, Russia). After PCR, the amount (%) of *BCR::ABL*/p210 mRNA was calculated according to the standard formula: the number of *BCR::ABL*/p210 copies was divided by the number of *ABL* copies and multiplied by 100.

DNA isolation

Genomic DNA was isolated from 0.2 mL of peripheral blood using a QIAamp DNA Mini Kit (Qiagen, USA) according to the manufacturer's protocol. DNA was eluted in 50 μL of the AE buffer (10 mM Tris-Cl, 0.5 mM EDTA, pH 9.0).

Screening for BCR::ABL tyrosine kinase domain mutations

The mutational analysis of BCR::ABL tyrosine kinase by Sanger sequencing was performed both after cDNA amplification preceding RT, and after amplification of a specific region of genomic DNA. In the first case, standard nested PCR was performed using oligonucleotides as well as amplification and thermocycling conditions as described previously [24]. In the second case, we performed long-range PCR using two-round PCR and a BioMaster LR HS-PCR reagent kit (Biolabmix, Russia) to identify mutations in the *BCR::ABL* oncogene in genomic DNA. Amplification and thermocycling conditions did not differ for both rounds of PCR, except for the use of specific oligonucleotides selected through the NCBI system (*Table 2*). Each sample for the second round of PCR contained 1× PCR buffer, 2.5 mM of each dNTP, 10 pM forward and reverse primers, 3 μL of the amplification

product (after the first round of PCR), and 5 units of Encyclo Taq polymerase (Evrogen, Russia). The thermocycling steps included the initial holding at 95°C for 10 min, followed by 50 cycles: 95°C for 15 s and 60°C for 1 min. The size of the amplification product after the second round of PCR was 293 bp. To determine the mutational status of *BCR::ABL*, Sanger sequencing was performed using an ABI PRISM 3500 genetic analyzer (Applied Biosystems, USA).

RESULTS

Mutation analysis at the onset of B-ALL

According to standard karyotyping, an abnormal karyotype was identified in the patient's bone marrow cells: {t(9;22)(q34;q11) [15], 46XY [5]}. A Ph chromosome was found in 15 of 20 mitoses (75%). Furthermore, the *BCR::ABL* oncogene was detected in 120 of the 200 interphase nuclei viewed using FISH. According to the quantitative PCR data, increased expression of the *BCR::ABL*/p210 oncogene was found in peripheral blood, its level not exceeding 56% (*Fig. 2*). No additional mutations were found in the *BCR::ABL* gene encoding the tyrosine kinase domain.

Therapy and monitoring of minimal residual disease (MRD)

After securing informed consent, treatment was initiated according to the Ph+ALL-2012m protocol, in combination with imatinib (600 mg) [25]. After 8 months of therapy that included two induction and three consolidation phases, complete hematologic remission and a profound molecular response (*BCR::ABL*/p210: 0.002%) were noted. However, six months later, the patient had a relapse (blasts: 15%, *BCR::ABL*/p210: 105%). After three weeks of therapy (100 mg dasatinib + 20 mg dexamethasone), the number of blasts was reduced to 4.8% and the level of *BCR::ABL*/p210 decreased to 11.8%. Sanger sequencing analysis showed a T-to-C nucleotide substitution at position 949 (NM_005157) of the *ABL* gene, resulting in the F317L point mutation. Meanwhile, a novel nine-nucleotide insertion (GCCCTTCCC) at a position between 1073 and 1074 (NM_005157) of the *ABL* gene was found. The latter mutation caused a lysine substitution at position K294 for four amino acid residues; namely, serine-proline-serine-glutamine (K294SPSQ), which preceded histidine at position H295 (*Fig. 3*).

Once the results of the *BCR::ABL* mutational status were available, dasatinib was replaced with bosutinib. Treatment with bosutinib (500 mg) + dexamethasone (40 mg), in combination with two courses of inotuzumab therapy (0.8 and 0.5 mg/m^2), was performed. The patient showed a decrease in the blast

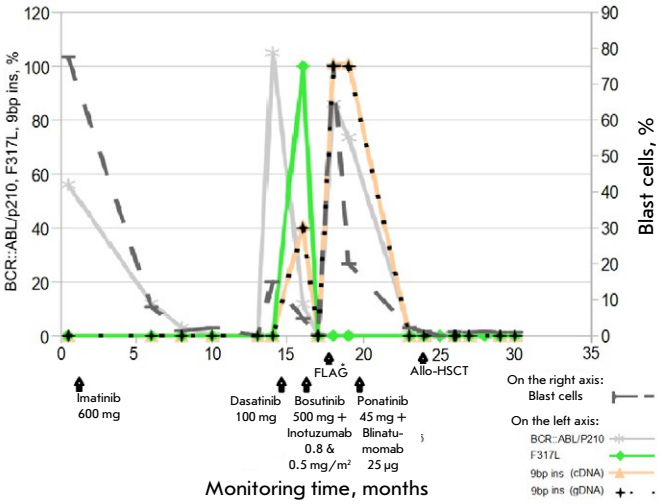


Fig. 2. Analysis of biomarkers at the onset of B-ALL and during treatment. FLAG* therapy includes FLAG + venetoclax (100 mg) + asciminib (400 mg). The number (%) of *BCR::ABL*/210 oncogene transcript mRNA (mutation) was estimated relative to the *ABL* reference gene (wild type). The number (%) of blast cells was determined relative to the total number of all nucleated cells in the bone marrow

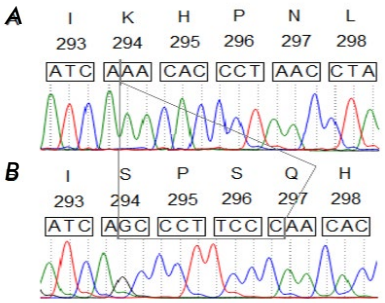


Fig. 3. The sequence of a novel insertion in the tyrosine kinase domain of BCR::ABL found in a B-ALL patient. (*A*) At the onset of B-ALL: no mutation was detected. (*B*) B-ALL relapse: insertion (K294SPSQ) was detected

count to 0.2% and the *BCR::ABL*/p210 mRNA expression level, to 0.069%. However, 1.5 months after completion of this therapy, an increase in the blast count to 75.2% and an increase in the *BCR::ABL*/p210 level to 86% were found in the patient's bone marrow. Sanger sequencing showed the absence of the F317L mutation but presence of a nine-nucleotide insertion. After FLAG+venetoclax (100 mg) + asciminib (400 mg) therapy, the blast count in the bone marrow decreased to 20% (*BCR::ABL*/p210: 73.5%). Sequencing

again confirmed the presence of the insertion. Change of therapy to ponatinib (45 mg) + blinatumomab (28 μg) resulted in the disappearance of the leukemia clone with the insertion, which correlated with a complete molecular response and complete clinical and hematologic remission. One month after chemotherapy, the patient underwent allo-HSCT from an available HLA-matched unrelated donor. Monitoring of the *BCR::ABL*/p210 oncogene expression and its mutational status confirmed the absence of any molecular genetic abnormalities during the last six months after allo-HSCT. The patient is currently in MRD-negative remission.

DISCUSSION

This article describes a rare clinical case of Ph-positive B-ALL with a chimeric variant of the *BCR::ABL* oncogene, typical of CML and characterized by a number of features. First, no mutations in the tyrosine kinase domain of BCR::ABL were detected in the patient at the onset of B-ALL (Ph+, *BCR::ABL*/p210+). Second, the development of resistance to dasatinib correlated with the detection of two tumor clones. One of them carried a point mutation F317L, while the other one carried a novel nine-nucleotide insertion accompanied by a four amino acid substitution of lysine at position K294 namely serine-proline-serine-glutamine (K294SPSQ). The insertion did not result in a reading frame shift. Third, after the emergence of resistance to dasatinib, the patient was found to be refractory to bosutinib and asciminib. Meanwhile, the disappearance of one leukemia clone carrying the F317L mutation and the presence of another clone carrying a nine-nucleotide insertion were noted. Only a therapy switch to ponatinib+blinatumomab resulted in complete eradication of the clone carrying the insertion. According to earlier literature data, the F317L point mutation localizes in the region of the BCR::ABL kinase domain that is responsible for imatinib binding (IM binding site) [26]. In addition, carriers of this mutation are resistant to imatinib and dasatinib but sensitive to bosutinib [18]. In our case, treatment (bosutinib+dexamethasone) in combination with two courses of inotuzumab therapy resulted in the disappearance of the F317L mutation. However, the nine-nucleotide insertion remained. Meanwhile, the patient had an elevated expression level of the *BCR::ABL*/p210 oncogene and increased number of tumor cells, which correlated with leukemia progression (*Fig. 2*). Interestingly, the nine-nucleotide insertion associated with the replacement of lysine by four new amino acid residues is located in the SH3 contact site of the tyrosine kinase domain of BCR::ABL. Mutations leading to resistance to TKIs are known to occur most frequently in this region as well as in the P-loop region [27]. According to earlier studies, the SH3 contact site is required for the autoinhibition of ABL tyrosine kinase in normal cells [28]. Mutations, and especially insertions at this site impair the aforementioned function. In addition, according to recent studies, mutations in the *BCR::ABL* gene leading to the modification of contact sites (SH2 and SH3), along with destabilization of specific tertiary structures of the protein and large-scale conformational changes, are considered to be additional mechanisms for the emergence of resistance to TKI [29]. In our case, the emergence of a nine-nucleotide insertion in the SH3 contact site of the tyrosine kinase domain of BCR::ABL led to the discovery of a new SPSQ motif. It appears that serine phosphorylation of this motif through the involvement of one of the serine/threonine kinases, such as Dyrk1A, may contribute to the progression of the leukemia. Dyrk1A was previously found to be involved in the phosphorylation of several targeting proteins, including Amph1, which also possesses an SPSQ motif [30]. Interestingly, increased expression of Dyrk1A was recently detected in B-ALL (*BCR::ABL*/p190+) [31]. Activation of the JAK/STAT-signaling pathway was shown to promote the proliferation of leukemia cells overexpressing Dyrk1A. The same investigators, using a mouse model of B-ALL (*BCR::ABL*/p190+), suggested that Dyrk1A may be involved in the regulation of *BCR::ABL* expression. In particular, an artificially induced heterozygous Dyrk1A deficiency in these mice may help prevent leukemic cell survival, thereby promoting the normalization of hematopoiesis.

CONCLUSIONS

In this study, a patient with B-ALL (*BCR::ABL*/p210+) developed resistance to several TKIs (dasatinib, busutinib, and asciminib), as well as to polychemotherapy including the FLAG regimen. This was accompanied by an increase in *BCR::ABL*/p210+ expression and the emergence of two leukemia clones, one of which included a point mutation (F317L) and the other one, a nine-nucleotide insertion (GCCCTTCCC) involving the substitution of lysine at position K294 for four new amino acid residues (K294SPSQ). In the described case of B-ALL, the elevated resistance to TKI may be due to an increase in the level of serine phosphorylation in a new SPSQ motif involving the serine/threonine kinase DyrkA1. This may lead to the activation of the JAK/STAT signaling pathway, ultimately enhancing cell proliferation and supporting leukemogenesis. ●

REFERENCES

1. Ribera J., Morgades M., Zamora L., Montesinos P., Gómez-Seguí I., Pratcorona M., Sarrà J., Guàrdia R., Nomdedeu J., Tormo M., et al. // Cancer. 2015. V. 121. № 21. P. 3809–3817. doi: 10.1002/cncr.29579.
2. Ishii Y., Nhiayi M.K., Tse E., Cheng J., Massimino M., Durden D.L., Vigneri P., Wang J.Y. // PLoS One. 2015. V. 10. № 10. P. e0140585. doi: 10.1371/journal.pone.0140585.
3. Manzella L., Tirro E., Pennisi M.S., Massimino M., Stella S., Romano C., Vitale S.R., Vigneri P. // Curr. Cancer Drug Targets. 2016. V. 16. № 7. P. 594–605. doi: 10.2174/1568009616666160105105857.
4. Ren R. // Nat. Rev. Cancer. 2005. V. 5. № 3. P. 172–183. doi: 10.1038/nrc1567.
5. Gleissner B., Gökbuget N., Bartram C.R., Janssen B., Rieder H., Janssen J.W., Fonatsch C., Heyll A., Voliotis D., Beck J., et al. // Blood. 2002. V. 99. № 5. P. 1536–1543. doi: 10.1182/blood.v99.5.1536.
6. Druker B.J., Sawyers C.L., Kantarjian H., Resta D.J., Reese S.F., Ford J.M., Capdeville R., Talpaz M. // N. Engl. J. Med. 2001. V. 344. № 14. P. 1038–1042. doi: 10.1056/NEJM200104053441402.
7. Chen J., Schmitt A., Chen B., Rojewski M., Rübeler V., Fei F., Yu Y., Yu X., Ringhoffer M., von Harsdorf S., et al. // J. Cell. Mol. Med. 2008. V. 12. № 5B. P. 2107–2118. doi: 10.1111/j.1582-4934.2008.00234.x.
8. Mahon F.X., Deininger M.W., Schultheis B., Chabrol J., Reiffers J., Goldman J.M., Melo J.V. // Blood. 2000. V. 96. № 3. 1070–1079.
9. Kantarjian H., Giles F., Wunderle L., Bhalla K., O'Brien S., Wassmann B., Tanaka C., Manley P., Rae P., Mietlowski W., et al. // N. Engl. J. Med. 2006. V. 354. № 24. P. 2542–2551. doi: 10.1056/NEJMoa055104.
10. Mishra S., Zhang B., Cunnick J.M., Heisterkamp N., Groffen J. // Cancer Res. 2006. V. 66. № 10. P. 5387–5393. doi: 10.1158/0008-5472.CAN-05-3058.
11. Travis J. // Science. 2004. V. 305. № 5682. P. 319–321. doi: 10.1126/science.305.5682.319a.
12. Wu J., Meng F., Kong L.Y., Peng Z., Ying Y., Bornmann W.G., Darnay B.G., Lamothe B., Sun H., Talpaz M, et al. // J. Natl. Cancer Inst. 2008. V. 100. № 13. P. 926–939. doi: 10.1093/jnci/djn188.
13. Rossari F., Minutolo F., Orciuolo E. // J. Hematol. Oncol. 2018. V. 11. № 1. P. 84. doi: 10.1186/s13045-018-0624-2.
14. Jabbour E., Kantarjian H., Cortes J. // Clin. Lymphoma Myeloma Leuk. 2015. V. 15. № 6. P. 323–334. doi: 10.1016/j.clml.2015.03.006.
15. Li H., Zhang W., Yi D., Ye Y., Xiao X. // Leuk. Lymphoma. 2017. V. 58. № 4. P. 1005–1007. doi: 10.1080/10428194.2016.1225205.
16. Ernst T., La Rosée P., Müller M.C., Hochhaus A. // Hematol. Oncol. Clin. North Am. 2011. V. 25. № 5. P. 997–1008. doi: 10.1016/j.hoc.2011.09.005.
17.Pfeifer H., Lange T., Wystub S., Wassmann B., Maier J., Binckebanck A., Giagounidis A., Stelljes M., Schmalzing M., Dührsen U., et al. // Leukemia. 2012. V. 26. № 7. P. 1475–1481. doi: 10.1038/leu.2012.5.
18. Leow B.C.S., Kok C.H., Yeung D.T., Hughes T.P., White D.L., Eadie L.N. // Sci. Rep. 2023. V. 13. № 1. P. 13110. doi: 10.1038/s41598-023-40279-2.
19. Kuang P., Liu T., Huang Q., Ye Y., Xiang B., Huang J., et al. // Leuk. Res. 2012. V. 36. № 8. P. e159–e162. doi: 10.1016/j.leukres.2012.04.019.
20. Hayette S., Chabane K., Tchirkov A., Berger M.G., Nicolini F.E., Tournilhac O., Diwu L., Wang Y., Meng W., Dong T., et al. // Haematologica. 2009. V. 94. № 9. P. 1324–1326. doi: 10.3324/haematol.2009.007864.
21.Kato K., Takagi S., Takano H., Tsunoda S., Watanabe O., Yamaguchi K., Kageyama K., Kaji D., Taya Y., Nishida A., et al. // Int. J. Hematol. 2024. V. 119. № 2. P. 205–209. doi: 10.1007/s12185-023-03691-y.
22. Verma R., Babu A. Human Chromosomes: Manual of Basic Techniques. New York: 1989. 240 p.
23. Bogdanov K.V., Nikulina T.S., Lomaia E.G., Slyadnev M.N., Zaritskey A.Y. // Russ. J. Bioorg. Chem. 2017. V. 43. № 5. P. 544–551. doi: 10.1134/S1068162017040033.
24. Branford S., Rudzki Z., Walsh S., Grigg A., Arthur C., Taylor K., Herrmann R., Lynch K.P., Hughes T.P. // Blood. 2002. V. 99. № 9. P. 3472–3475. doi: 10.1182/blood.v99.9.3472.
25.Savchenko V.G., Parovichnikova E.N., Afanasiev B.V., Troitskaya V.V., Alyoshina O.A., Sokolov A.N., Kuzmina L.A., Klyasova G.A., Bondarenko S.N., Kaplanov K.D., et al. Clinical Recommendations on Diagnosis and Treatment of Acute Lymphoblastic Leukemia in Adults. M.: National Hematology Society, 2018. 110 p.
26. O'Hare T., Eide C.A., Deininger M.W. // Blood. 2007. V. 110. № 7. P. 2242–2249. doi: 10.1182/blood-2007-03-066936.
27. Wongboonma W., Thongnoppakhun W., Auewarakul C.U. // Exp. Mol. Pathol. 2012. V. 92. № 2. P. 259–265. doi: 10.1016/j.yexmp.2012.01.007.
28. Chomel J.C., Sorel N., Turhan A.G. // Stem Cells and Cancer Stem Cells. Ed. Hayat M. Dordrecht, Germany: Springer, 2012. 315 p.
29. Azam M., Latek R.R., Daley G.Q. // Cell. 2003. V. 112. № 6. P. 831–843. doi: 10.1016/s0092-8674(03)00190-9.
30. Katayama S., Sueyoshi N., Kameshita I. // Biochemistry. 2015. V. 54. № 19. P. 2975–2987. doi: 10.1021/bi501308k.
31.Bhansali R.S., Rammohan M., Lee P., Laurent A.P., Wen Q., Suraneni P., et al. // J. Clin. Invest. 2021. V. 131. № 1. P. E135937. doi: 10.1172/JCI135937.

The Drosophila Zinc Finger Protein Aef1 Colocalizes with Enhancers and Is Involved in the Transcriptional Regulation of Numerous Genes

N. E. Vorobyeva[1], J. V. Nikolenko[2], A. N. Krasnov[1*]
[1]Institute of Gene Biology, Russian Academy of Sciences, Moscow, 119334 Russia
[2]Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia
*E-mail: krasnov@genebiology.ru
Received November 07, 2024; in final form, March 06, 2025
DOI: 10.32607/actanaturae.27556

ABSTRACT **In our previous studies, we demonstrated that the Drosophila zinc finger protein Aef1 interacts with the SAGA DUB module. The Aef1 binding sites colocalize with the SAGA histone acetyltransferase complex and the dSWI/SNF chromatin remodeling complex, as well as the origin recognition complex (ORC). Aef1 predominantly localizes with the promoters of active genes (55% of all sites) and can be involved in transcriptional regulation. In this study, we showed that Aef1 binding sites in Drosophila S2 cells, located outside gene promoters, are nucleosome-depleted regions and colocalize with the SAGA, dSWI/SNF, and ORC complexes. Aef1 binding sites colocalize with the CBP protein and the H3K27Ac histone tag, which is considered to be an active enhancer mark. An RNA-Seq experiment was conducted in Drosophila S2 cells, both normal and with RNA interference targeting the Aef1 protein, to study the role played by the Aef1 protein in transcriptional regulation. The Aef1 protein was shown to affect the transcription of 342 genes, more than half of those (178 genes) containing Aef1 at their promoters or enhancers. Hence, we infer that the Aef1 protein is recruited to both promoters and enhancers and is involved, both directly and indirectly, in the regulation of the transcription of the respective genes.**
KEYWORDS **Aef1, SAGA, dSWI/SNF, ORC, CBP, H3K27Ac, enhancers.**
ABBREVIATIONS **SAGA – histone acetyltransferase complex; SWI/SNF – chromatin remodeling complex; ORC – origin recognition complex.**

INTRODUCTION

Regulation of eukaryotic gene expression is a complex process involving several successive stages of transcription, mRNA processing, mRNA export from the nucleus, translation, and protein folding [1]. Local chromatin structure, gene location relative to functional nuclear compartments, and long-range interactions between cell regulatory elements constitute an additional level in the regulation of genetic processes in the context of the complex architecture of the eukaryotic genome in the 3D space of the cell nucleus [2–5].

The highly conserved SAGA coactivator complex, chromatin histone modification (acetylation and deubiquitination) being its main function, is composed of more than 20 protein subunits [6]. The SAGA complex subunits interact with various transcriptional activators, thereby recruiting the complex to the promoters of specific genes [7, 8]. There is a degree of synergism between the chromatin remodeling complex and chromatin modifying complex. The SAGA complex has been shown to acetylate the nucleosomes on gene promoters during transcription activation, leading to the recruitment of the dSWI/SNF chromatin remodeling complex and stimulation of its remodeling activity [9, 10]. The SWI/SNF chromatin remodeling complex and CBP/p300/Nejire acetyltransferase, which is responsible for tagging active enhancers with H3K27Ac, exhibit considerable functional action on the activation of the ecdysone-dependent genes *dhr3* and *hr4* in S2 cells [11–14]. Histone H3 acetylation (H3K27Ac) at enhancers was shown to be required for the activation of ecdysone-dependent genes [15].

The Drosophila SAGA and SWI/SNF complexes reside within various regulatory elements of the genome, including promoters, where they often colocalize with the origin recognition complex (ORC) [16]. Replication is initiated at numerous sites known as replication origin sites. The ORC composed of six subunits (ORC1–6) is recruited to the replication origins. The ORC binds specific genomic regions [17]; however, the subunits of this complex exhibit no explicit DNA sequence specificity. Therefore, a question arises: which factors are responsible for the positioning of ORCs in the genome?

Our previous studies revealed that the insulator protein Su(Hw) carrying zinc finger domains interacts with the ENY2 protein (a subunit of the SAGA complex) and recruits the SAGA, SWI/SNF, and ORC complexes to Su(Hw)-dependent Drosophila insulators, thus being simultaneously involved in transcriptional regulation and the positioning of replication origins [18–22]. A hypothesis has been put forward that there also are other proteins carrying zinc finger domains which interact with the Drosophila SAGA complex and function in a similar manner at other regulatory elements of the genome, including promoters. Further experiments have identified four additional proteins having zinc finger domains: CG9890, CG9609, Aef1 (Adult enhancer factor 1), and CG10543 [23–27]. These proteins also colocalize with the SAGA, ORC, and dSWI/SNF complexes at their binding sites, preferentially at active gene promoters' sites, and can be involved in transcriptional regulation. As shown previously, the Aef1 protein is recruited to the enhancers of the *adh*, *yp1*, and *yp2* genes and affects their transcription [28–30]. Our earlier study [26] revealed that RNA interference targeting the Aef1 protein affects the transcription of several genes. In order to assess the impact of Aef1 on the transcription of all the genes in Drosophila S2 cells, we conducted an RNA-Seq experiment both under normal conditions and upon RNA interference targeting the Aef1 protein.

MATERIALS AND METHODS

Cultivation of Schneider 2 (S2) cells. RNA interference

S2 cells were cultured in Schneider's Insect Medium (Sigma, USA) supplemented with 10% fetal bovine serum (HyClone, USA) at 25°C. The cells were transfected using the Effectene Transfection Reagent (Qiagen, USA), according to the manufacturer's protocol. Knockdown of the *Aef1* gene was performed via RNA interference according to a published protocol [22]. The dsRNA corresponding to a fragment of plasmid pBluescript II SK(-) (Stratagene, USA) was used as nonspecific control for RNA interference. The dsRNA for knockdown of the *Aef1* gene and control was synthesized using the following primers: Aef1, GAATTAATACGACTCACTATAGGGAGAATGATGCATATCAAAAGCCT and GAATTAATACGACTCACTATAGGGAGATCCGGGATGCTCGCTATGT; pBluesciptIISK(-), GAATTAATACGACTCACTATAGGGAGAGTTACATGATCCCCCATG and GAATTAATACGACTCACTATAGGGAGATTTCGCCCCGAAGAACG.

For each RNA interference experiment, 30 μg dsRNA per 10^6 cells was used. The experiment was conducted in three replicates. RNA was extracted after 5-day incubation.

RNA-Seq and identification of differentially expressed genes

RNA-Seq libraries were constructed using a NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs). The library quality was verified using Bioanalyzer. The libraries were sequenced on an Illumina HiSeq 2000 genome sequencing system. Raw reads in the Fastq format were aligned to the Drosophila genome dmel_r6.54 using the Hisat2 software [31]; adapters had preliminarily been removed in the Atropos software [32]. The "-a" key enabling the search for multiple alignments to be excluded from the analysis was also employed. Only the unique mapped reads were used for further work by analyzing the NH:i tag among the output data in the Hisat2 software. Differentially expressed genes were identified in the CuffDiff2 software [33].

Enrichment analysis of protein factors at Aef1 binding sites

The ChIP-Seq profiles of the Aef1, GCN5, OSA, ORC2, H3, CBP, and H3K27Ac proteins obtained earlier [14, 24, 26, 27, 34] were used to study the colocalization of Aef1 binding sites with various protein factors. This analysis and data visualization were performed using the deepTool2 suite [35].

RESULTS AND DISCUSSION

The Aef1 binding sites are nucleosome-depleted regions and colocalize with the SAGA, dSWI/SNF, and ORC complexes regardless of the genomic localization

We have previously demonstrated that the Aef1 protein predominantly resides at the promoters of active genes (55% of sites) and is involved in transcriptional regulation [26]. Aef1 binding sites colocalize with the chromatin modification and chromatin remodeling complexes SAGA and dSWI/SNF, as well as with the

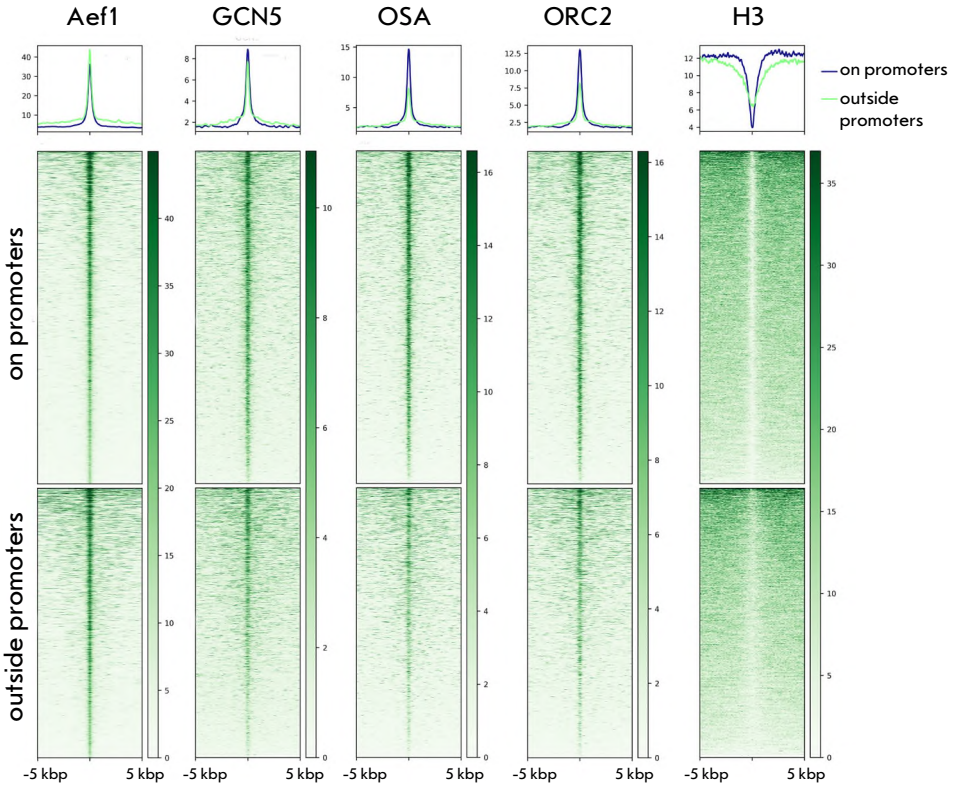


Fig. 1. Genomic colocalizations of Aef1 binding sites with GCN5 (the SAGA complex), OSA (the dSWI/SNF complex), ORC2 (the ORC complex), and histone H3. The names of the respective proteins are displayed at the top of the panels. Data are shown for Aef1 sites located in promoters (the middle panel) and outside promoters (the lower panel). The upper panel displays averaged profiles. The blue line represents the profile of proteins located at Aef1 promoter sites, while the green line represents the profile of the proteins on Aef1 sites outside promoters

ORC replication complex. A considerable portion of Aef1 sites (35%) are located within gene bodies (excluding promoters) and intergenic regions (10%). This study focuses on the properties of the binding sites residing outside promoters. The ChIP-Seq profiles of the proteins Aef1, GCN5 (the SAGA complex), OSA (the SWI/SNF complex), ORC2 (the ORC complex), and histone H3 were utilized for the analysis [14, 24, 26, 27]. We analyzed the enrichment in each of these proteins at two categories of sites: the Aef1 binding sites at promoters and outside them (*Fig. 1*). The results show that the protein complexes under study are recruited to both groups of sites with approximately equal efficiency, although the sites outside the promoters are characterized by lower levels of the OSA and ORC2 proteins. An analysis of the histone H3 distribution revealed that all the Aef1 binding sites were nucleosome-depleted regions, which is typical of active regulatory elements involved in transcriptional regulation [36].

Aef1 binding sites colocalize with active enhancers

The Aef1 binding sites located outside promoters were analyzed to better understand the nature of the binding sites. Several studies have shown that Aef1 is recruited to enhancers of the *adh*, *yp1*, and *yp2* genes, affecting their transcription [28–30]. Histone tag H3K27Ac mediated by acetyltransferase CBP/p300/Nejire is an active enhancer mark. The chromatin remodeling complex SWI/SNF and acetyltransferase CBP/p300/Nejire are recruited to ecdysone-dependent enhancers, which is needed for transcription activation [14]. We aimed to examine the genomic colocalization of the CBP protein and histone tag H3K27Ac at Aef1 binding sites. The previously obtained ChIP-Seq profiles of CBP and H3K27Ac [14, 34] were used. *Figure 2* demonstrates that Aef1 binding sites are characterized by CBP and H3K27Ac recruitment. Therefore, it can be inferred that a significant portion of Aef1 binding sites colocalize with active enhancers.

We noticed that CBP and H3K27Ac also tag the Aef1 binding sites residing in promoters and analyzed another group of sites; namely, the promoters lacking the Aef1 protein (*Fig. 2*, lower panels). An analysis of these sites revealed low CBP and H3K27Ac levels. Hence, recruitment of CBP and the high-intensity H3K27Ac signal at Aef1-carrying promoters correlate with the presence of an Aef1 binding site rather than with the promoter in general.

We decided to identify the potential consensus motif responsible for Aef1 binding at promoters and outside them using the MEME-ChIP software. An identical consensus motif (CAA)n was identified in

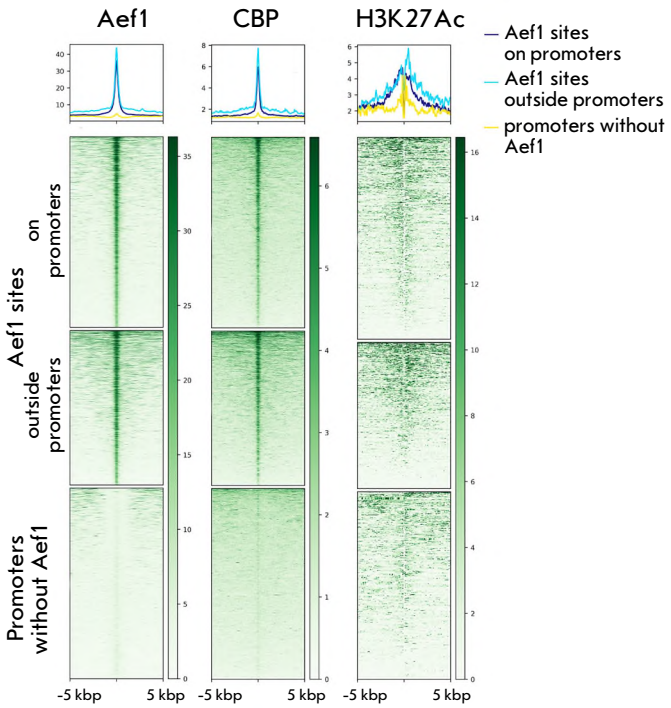


Fig. 2. Genomic colocalizations of the Aef1, CBP, and H3K27Ac proteins at three groups of sites: at Aef1 sites on promoters (upper panels); at Aef1 sites outside promoters (middle panels); and at promoters without Aef1 (lower panels). The names of the respective proteins are displayed at the top of the panels

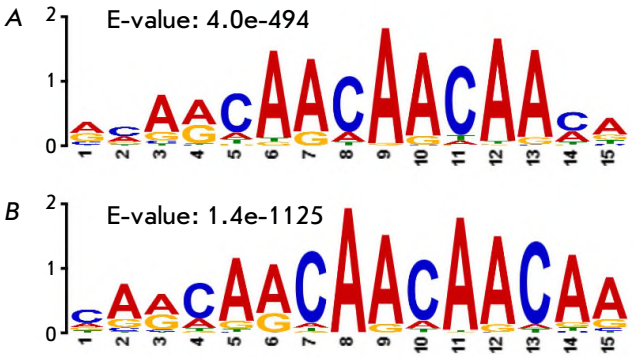


Fig. 3. The potential consensus binding motif of the Aef1 protein identified at promoter sites (*A*) and outside promoters (*B*). The E-value shows the statistical significance of the result and represents the probability of a random match

both groups of sites (*Fig. 3*), which had been identified previously across the entire set of sites [26], suggesting that both groups of sites appear owing to the DNA-binding properties of Aef1 *per se* rather than via looping of other regulatory elements. These findings are consistent with the data showing that an experimentally confirmed Aef1 binding site within the *adh* gene enhancer contains the CAACAA sequence.

The Aef1 protein is involved in transcriptional regulation

As mentioned earlier, Aef1 binding sites are localized both within and outside gene promoter regions. Both groups of sites colocalize with active enhancer marks, indicating that the Aef1 protein could play a role in transcriptional regulation. For this purpose, we conducted a RNA-Seq analysis on Drosophila S2 cells (under normal conditions and with RNA interference targeting the Aef1 protein). The analysis identified 342 genes whose expression was significantly altered (q-value < 0.05) upon RNA interference targeting Aef1. All the genes affected by RNA interference targeting the Aef1 protein were categorized into several groups, depending on whether there were Aef1 binding sites within the gene. It turned out that 57 (17%) genes contained Aef1 exclusively in their promoters, 52 (15%) genes contained Aef1 only in potential enhancers, and 69 (20%) genes had Aef1 in both promoters and enhancers. A total of 164 (48%) genes lacked Aef1 binding sites. These findings suggest that the Aef1 protein localized in both promoters and enhancers is involved in transcriptional regulation. The results also imply that Aef1 may act either directly or indirectly, since there are no binding sites in half of the genes. It is fair to assume that there may exist looping between Aef1 binding sites (potential enhancers) and distal promoters.

CONCLUSIONS

This study has demonstrated that the Aef1 zinc finger protein is involved in transcriptional regulation. RNA interference targeting the Aef1 protein affects the transcription of 342 genes in Drosophila S2 cells. Approximately half of these genes carry Aef1 binding sites in neither the promoter region nor the gene's body, which may be indicative of the indirect mechanisms of transcriptional regulation (e.g., via looping between enhancers and promoters). An analysis of Aef1 binding sites demonstrated that they colocalize with active enhancer marks: CBP protein and histone tag H3K27Ac. It is the general property of Aef1 binding sites that is independent of whether they reside in promoters or in the intergenic regions. Aef1-carrying promoters are much more enriched in the CBP protein and H3K27Ac compared to promoters lacking Aef1. That suggests that this property is specific to Aef1 binding sites rather than to promoters in general. It is known that several Drosophila enhancers containing Aef1 reside near transcription start sites (e.g.,

the *adh* gene enhancer) [30]. It can be hypothesized that Aef1 is a purely enhancer-associated protein, and that its localization at promoters may result from its recruitment to adjacent enhancers.

Our earlier studies [18, 19] have demonstrated that the Su(Hw) protein recruits the SAGA and dSWI/SNF complexes to its binding sites, resulting in the formation of nucleosome-depleted regions and recruitment of the ORC replication complex. Further experiments identified four additional zinc finger proteins (CG9890, CG9609, Aef1, and CG10543) colocalized with the SAGA, SWI/SNF, and ORC complexes [23–27]. The Su(Hw) protein is predominantly located in intergenic regions at insulators, while the CG9890, CG9609, and CG10543 proteins mainly localize with promoters. Our study has demonstrated that Aef1 binding sites colocalize with active enhancer marks. Despite the differences in genomic localization, all these proteins share properties with respect to the SAGA, SWI/SNF, and ORC complexes. We hypothesize that positioning of the ORC complexes in the genome is regulated by the DNA-binding proteins responsible for the formation of various regulatory elements, including insulators, promoters, and enhancers. We have demonstrated that Aef1 can be an example of such a protein. ●

This work was supported by the Russian Science Foundation (grant No. 20-14-00269).

REFERENCES

1. Orphanides G., Reinberg D. // Cell. 2002. V. 108. № 4. P. 439–451. doi: 10.1016/s0092-8674(02)00655-4.
2. Maksimenko O., Georgiev P. // Front Genet. 2014. V. 5. P. 28. doi: 10.3389/fgene.2014.00028.
3. van Bemmel J.G., Pagie L., Braunschweig U., Brugman W., Meuleman W., Kerkhoven R.M., van Steensel B. // PLoS One. 2010. V. 5. № 11. P. e15013. doi: 10.1371/journal.pone.0015013.
4. Rando O.J., Chang H.Y. // Annu Rev. Biochem. 2009. V. 78. P. 245–271. doi: 10.1146/annurev.biochem.78.071107.134639.
5. Tchurikov N.A., Krasnov A.N., Ponomarenko N.A., Golova Y.B., Chernov B.K. // Nucleic Acids Res. 1998. V. 26. № 13. P. 3221–3227. doi: 10.1093/nar/26.13.3221.
6. Koutelou E., Hirsch C.L., Dent S.Y. // Curr. Opin Cell Biol. 2010. V. 22. № 3. P. 374–382. doi: 10.1016/j.ceb.2010.03.005.
7. Baker S.P., Grant P.A. // Oncogene. 2007. V. 26. № 37. P. 5329–5340. doi: 10.1038/sj.onc.1210603.
8. Brown C.E., Howe L., Sousa K., Alley S.C., Carrozza M.J., Tan S., Workman J.L. // Science. 2001. V. 292. № 5525. P. 2333–2337. doi: 10.1126/science.1060214.
9. Chatterjee N., Sinha D., Lemma-Dechassa M., Tan S., Shogren-Knaak M.A., Bartholomew B. // Nucleic Acids Res. 2011. V. 39. № 19. P. 8378–8391. doi: 10.1093/nar/gkr535.
10. Li B., Carey M., Workman J.L. // Cell. 2007. V. 128. № 4. P. 707–719. doi: 10.1016/j.cell.2007.01.015.
11. Mazina M.Y., Kovalenko E.V., Derevyanko P.K., Nikolenko J.V., Krasnov A.N., Vorobyeva N.E. // Biochim. Biophys. Acta. 2018. V. 1861. № 2. P. 178–189. doi: 10.1016/j.bbagrm.2018.01.016.
12. Mazina M.Y., Nikolenko J.V., Fursova N.A., Nedil'ko P.N., Krasnov A.N., Vorobyeva N.E. // Cell Cycle. 2015. V. 14. № 22. P. 3593–3601. doi: 10.1080/15384101.2015.1100772.
13. Mazina M.Y., Kocheryzhkina E.V., Nikolenko J.V., Krasnov A.N., Georgieva S.G., Vorobyeva N.E. // Dokl. Biochem. Biophys. 2017. V. 473. № 1. P. 145–147. doi: 10.1134/S1607672917020144.
14. Krasnov A.N., Evdokimova A.A., Mazina M.Y., Erokhin M., Chetverina D., Vorobyeva N.E. // Int. J. Mol. Sci. 2023. V. 24. № 14. P. 11844. doi: 10.3390/ijms241411844.
15. Cheng D., Dong Z., Lin P., Shen G., Xia Q. // Int. J. Mol. Sci. 2022. V. 23. № 18. P. 10791. doi: 10.3390/ijms231810791.
16. MacAlpine D.M., Rodriguez H.K., Bell S.P. // Genes Dev. 2004. V. 18. № 24. P. 3094–3105. doi: 10.1101/gad.1246404.
17. Eaton M.L., Prinz J.A., MacAlpine H.K., Tretyakov G., Kharchenko P.V., MacAlpine D.M. // Genome Res. 2011. V. 21. № 2. P. 164–174. doi: 10.1101/gr.116038.110.
18. Vorobyeva N.E., Mazina M.U., Golovnin A.K., Kopytova D.V., Gurskiy D.Y., Nabirochkina E.N., Georgieva S.G., Georgiev P.G., Krasnov A.N. // Nucleic Acids Res. 2013. V. 41. № 11. P. 5717–5730. doi: 10.1093/nar/gkt297.
19. Mazina M., Vorob'eva N.E., Krasnov A.N. // Tsitologiia. 2013. V. 55. № 4. P. 218–224. doi:
20. Kurshakova M., Maksimenko O., Golovnin A., Pulina M., Georgieva S., Georgiev P., Krasnov A. // Mol. Cell. 2007. V. 27. № 2. P. 332–338. doi: 10.1016/j.molcel.2007.05.035.
21. Vorobyeva N.E., Erokhin M., Chetverina D., Krasnov A.N., Mazina M.Y. // Sci. Rep. 2021. V. 11. № 1. P. 16963. doi: 10.1038/s41598-021-96488-0.
22. Vorobyeva N.E., Krasnov A.N., Erokhin M., Chetverina D., Mazina M. // Epigenetics Chromatin. 2024. V. 17. № 1. P. 17. doi: 10.1186/s13072-024-00541-x.
23. Nikolenko J.V., Fursova N.A., Mazina M.Y., Vorobyeva N.E., Krasnov A.N. // Mol. Biol. (Moskow). 2022. V. 56. № 4. P. 557–563. doi: 10.31857/S0026898422040085.
24. Fursova N.A., Mazina M.Y., Nikolenko J.V., Vorobyova N.E., Krasnov A.N. // Acta Naturae. 2020. V. 12. № 4. P. 114–119. doi: 10.32607/actanaturae.11056.
25. Fursova N.A., Nikolenko J.V., Soshnikova N.V., Mazina M.Y., Vorobyova N.E., Krasnov A.N. // Acta Naturae. 2018. V. 10. № 4. P. 110–114. doi: 10.32607/20758251-2018-10-4-110-114.
26. Nikolenko J.V., Kurshakova M.M., Kopytova D.V., Vdovina Y.A., Vorobyova N.E., Krasnov A.N. // Mol. Biol. (Moskow). 2024. V. 58. № 4. P. 700–707. doi: 10.1134/S0026893324700286.
27. Nikolenko J.V., Kurshakova M.M., Kopytova D.V., Vdovina Y.A., Vorobyova N.E., Krasnov A.N. // Mol. Biol. (Moskow). 2024. V. 58. № 4. P. 693–699. doi: 10.1134/S0026893324700274.
28. An W., Wensink P.C. // Genes Dev. 1995. V. 9. № 2. P. 256–266. doi: 10.1101/gad.9.2.256.

29. Brodu V., Mugat B., Fichelson P., Lepesant J.A., Antoniewski C. // Development. 2001. V. 128. № 13. P. 2593–2602. doi: 10.1242/dev.128.13.2593.
30. Falb D., Maniatis T. // Genes Dev. 1992. V. 6. № 3. P. 454–465. doi: 10.1101/gad.6.3.454.
31. Kim D., Paggi J.M., Park C., Bennett C., Salzberg S.L. // Nat. Biotechnol. 2019. V. 37. № 8. P. 907–915. doi: 10.1038/s41587-019-0201-4.
32. Didion J.P., Martin M., Collins F.S. // Peer. J. 2017. V. 5. P. e3720. doi: 10.7717/peerj.3720.
33. Trapnell C., Hendrickson D.G., Sauvageau M., Goff L., Rinn J.L., Pachter L. // Nat. Biotechnol. 2013. V. 31. № 1. P. 46–53. doi: 10.1038/nbt.2450.
34. Mazina M.Y., Kovalenko E.V., Vorobyeva N.E. // Sci. Rep. 2021. V. 11. № 1. P. 172. doi: 10.1038/s41598-020-80650-1.
35. Ramirez F., Ryan D.P., Gruning B., Bhardwaj V., Kilpert F., Richter A.S., Heyne S., Dundar F., Manke T. // Nucleic Acids Res. 2016. V. 44. № W1. P. W160–165. doi: 10.1093/nar/gkw257.
36. McKay D.J., Lieb J.D. // Dev. Cell. 2013. V. 27. № 3. P. 306–318. doi: 10.1016/j.devcel.2013.10.009.

Cis-regulatory Function of the *Pou5f1* Gene Promoter in the Mouse MHC Locus

V. V. Ermakova#, E. V. Aleksandrova#, A. A. Kuzmin*, A. N. Tomilin**
Institute of Cytology, Russian Academy of Sciences, St. Petersburg, 194064 Russia
#contributed equally
*E-mail: a.kuzmin@incras.ru; **E-mail: a.tomilin@incras.ru
Received December 12, 2024; in final form, February 28, 2025
DOI: 10.32607/actanaturae.27596

ABSTRACT **The *Pou5f1* gene encodes the Oct4 protein, one of the key transcription factors required for maintaining the pluripotent state of epiblast cells and the viability of germ cells. However, functional genetics provides convincing evidence that *Pou5f1* has a broader range of functions in mouse ontogeny, including suppression of atherosclerotic processes. Related studies have primarily focused on the functions of the Oct4 protein, while the regulatory sequences within the *Pou5f1* gene have not been considered. In this study, we have developed a genetic model which is based on mouse embryonic stem cells (ESCs) for assessing the roles of the *Pou5f1* gene promoter in the transcriptional regulation of neighboring genes within the major histocompatibility complex (MHC) locus. We have demonstrated that deletion of this promoter affects the expression of selected genes within this locus neither in ESCs nor in the trophoblast derivatives of these cells. A notable exception is the *Tcf19* gene, which is upregulated upon *Pou5f1* promoter deletion and might be associated with the atherosclerosis pathology due to its pro-inflammatory activity. The developed genetic model will pave the way for future studies into the functional contribution of the *cis*-regulatory association of *Pou5f1*, *Tcf19*, and, possibly, other genes with the atherosclerotic phenotype previously reported for mice carrying the *Pou5f1* promoter deletion in vascular endothelial and smooth muscle cells.**
KEYWORDS ***Pou5f1*, Oct4, embryonic stem cells (ESCs), major histocompatibility complex, trophectoderm, regulation of gene expression.**
ABBREVIATIONS **iPSC – induced pluripotent stem cells; TLCs – trophoblast-like cells; ESCs – embryonic stem cells; MEFs – mouse embryonic fibroblasts; MMC – mitomycin C; Fgf4 – fibroblast growth factor 4; IFNγ – interferon γ; LPS – lipopolysaccharide; MHC – major histocompatibility complex; gRNA – guide RNA; GR – glucocorticoid receptor.**

INTRODUCTION

The Oct4 protein, which is also known as a component of the Yamanaka cocktail and is used for the reprogramming of somatic cells into induced pluripotent stem cells (iPSCs), is among the key factors responsible for maintaining the pluripotent state of epiblast cells and their cultured analogs, embryonic stem cells (ESCs) [1]. ESCs and iPSCs, collectively referred to as pluripotent stem cells (PSCs), are capable of unlimited proliferation and differentiation into any type of somatic cells. The aforementioned properties make these cells a valuable tool for studying early embryogenesis, *in vitro* modeling of genetic diseases, and developing approaches in regenerative medicine. The self-maintenance and the choice of differentiation lineage of PSCs critically depend on Oct4 expression [2], with even slight changes in its levels having a significant effect on the fate of the PSCs [3, 4].

The transcription factor Oct4 is encoded by the *Pou5f1* gene, which resides within the major histocompatibility complex (MHC) gene cluster. The *Pou5f1* gene is located on the short arm of human chromosome 6 and on mouse chromosome 17 (*Fig. 1*). In both cases, this locus is among the most densely packed genomic regions [5] and comprises numerous genes encoding the proteins involved in the innate and adaptive immune responses and, particularly, those responsible for antigen processing and presentation [6].

Until today, it has been believed that a distal enhancer interacting with the *Pou5f1* promoter in "naïve" PSCs, as well as a proximal enhancer being active in primed pluripotent cells, are sufficient to provide for the regulation of *Pou5f1* expression and, therefore, proper functioning of PSCs and their proper exit from pluripotency [7, 8]. However, along with

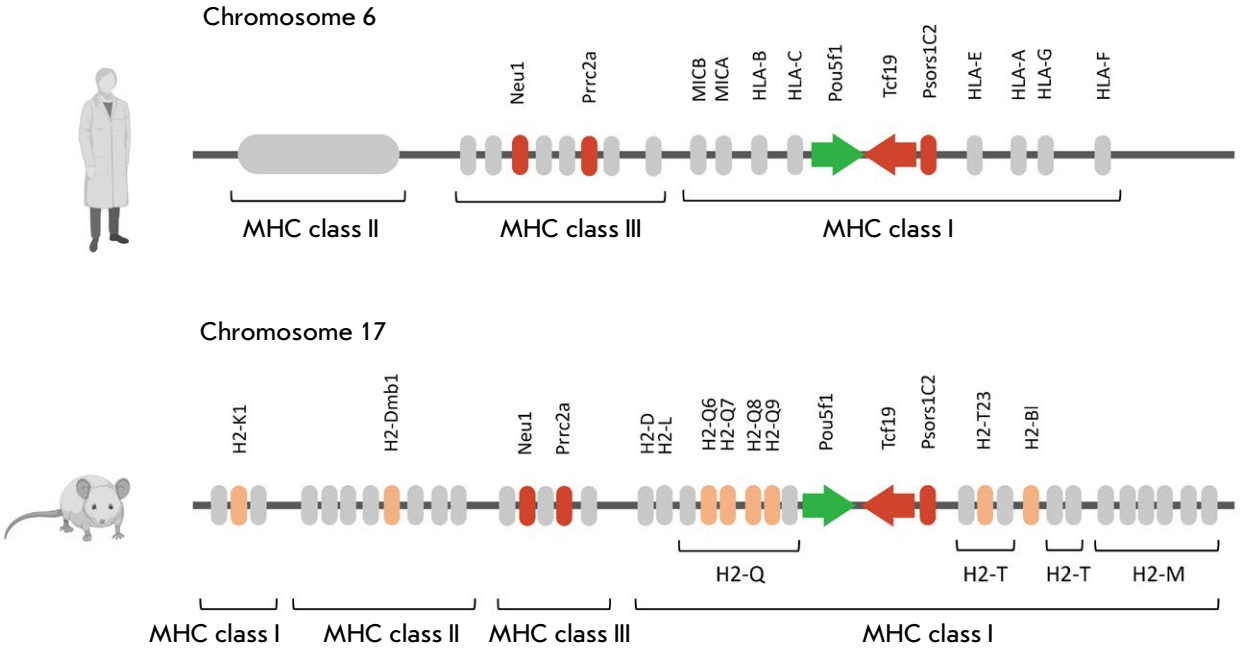


Fig. 1. Schematic representation of the *Pou5f1-MHC* locus. A schematic depiction of the *Pou5f1-MHC* locus for human (top) and mouse (bottom). Genes analyzed in this study are highlighted: *Pou5f1*, in green; MHC genes, in orange; the genes potentially interacting with *Pou5f1*, including *Tcf19*, in red. The directions of transcription of the *Pou5f1* and *Tcf19* genes are additionally indicated with arrows. The figure was created using BioRender

the classical regulatory elements of the *Pou5f1* gene (the promoter, distal and proximal enhancers) described by Yeom et al. back in 1996 [9], advances in high-throughput sequencing techniques have led to the discovery of numerous, previously unknown *cis*-regulatory elements that affect the expression of this gene [10, 11]. Hence, it has become clear that regulation of the *Pou5f1* gene is a much more fine-tuned process than previously thought. To date, the specific roles of the individual regulatory elements involved in *Pou5f1* expression control have been poorly characterized. Diao et al. demonstrated that just 17 out of the 41 identified regulatory elements of *Pou5f1* serve as promoters for other protein-coding genes, including its nearest neighbor – *Tcf19* [10]; however, it is unclear whether there is an opposite *cis*-regulatory association between *Pou5f1* and the neighboring genes. Some findings showing a correlation between the risk of developing psoriasis and polymorphisms in the promoter region and the first exon of the *Pou5f1* gene imply that there can be such an association [12].

An inverse correlation between *Pou5f1* and *MHC* gene expression during ontogenesis has an interesting aspect. It is believed that in mouse ESCs, the expression level of MHC class I and II genes is low, while it increases during the differentiation of these cells [13, 14]. Meanwhile, according to the current paradigm, *Pou5f1* expression is confined to PSCs and germ cells [9]. Therefore, it is possible that the protein-encoding activity of the *Pou5f1* gene switches to the *cis*-regulatory one required to activate *MHC* genes. This mechanism is consistent with the findings in experiments on mice carrying a deletion of the *Pou5f1* promoter region in smooth muscle and endothelial cells, which have shown a significantly deteriorated atherosclerotic phenotype, causing reduced plaque stability, lipid accumulation, inflammation, reduction of the mitochondrial membrane potential in endothelial cells, and decreased smooth muscle cell migration [15, 16].

In this study, we developed a genetic model that allowed us to assess the *cis*-regulatory function of the *Pou5f1* promoter region with respect to the genes within the *Pou5f1-MHC* locus in ESCs and their differentiated progeny. Following a successful differentiation of ESCs into the trophoblast lineage via forced Cdx2 expression, we did not observe any regulatory role of the *Pou5f1* promoter region in the expression of various genes within the MHC locus. However, we found that the *Pou5f1* promoter represses the expression of the *Tcf19* gene in both mouse ESCs and their trophoblastic derivatives.

EXPERIMENTAL

Obtaining mitotically inactivated embryonic fibroblasts

Mouse embryonic fibroblasts (MEFs) were isolated in accordance with the current animal welfare laws of the Russian Federation, with approval from the Institute's Ethics Committee (protocol No. 12/23).

MEFs derived from C57BL/6 mouse embryos (12-14 d.p.c.) were cultured on adhesive plastic pre-treated with a 0.1% gelatin solution (Sigma, USA). The cells were cultured in a DMEM GlutaMAX medium (Gibco, USA) supplemented with 10% HyClone FBS (Cytiva, USA) and 1× penicillin/streptomycin (Gibco). After 4–5 passages, once a confluent cell monolayer had been formed, the MEFs were incubated for 2.5 h in a medium supplemented with 10 μg/mL mitomycin C (MMC, Sigma). After incubation, the cells were washed with PBS and cryopreserved for future use.

Culturing of ESCs

Mouse embryonic stem cells (ESCs) were cultured at 37°C in a humidified atmosphere containing 5% CO_2 on plates for adherent cell cultures. A feeder layer of mitotically inactivated mouse embryonic fibroblasts (MMC-MEFs) with a density of 36 × 10^3 cells/cm^2, seeded into wells one day prior to the addition of ESCs, was used as a substrate. The cells were cultured in a standard S/L ESC medium containing KnockOut DMEM (Gibco) supplemented with 15% HyClone FBS (Cytiva), 1×NEAA (Gibco), 1× penicillin/streptomycin (Gibco), 0.1 mM β-mercaptoethanol (Sigma-Aldrich), 2 mM L-glutamine (Gibco), and 1 : 5,000 in-house generated hLIF.

For reverting ESCs to the naïve pluripotent state, we used the 2i/L medium containing N2B27 (a mixture of DMEM/F12 (Gibco) and Neurobasal (1 : 1)) enriched with 1× N2, 1× B27 (without retinoic acid, Gibco), 50 μM β-mercaptoethanol (Sigma-Aldrich), 0.005% BSA (Sigma), 1× penicillin/streptomycin (Gibco), and 2 mM L-glutamine (Gibco) supplemented with 3 μM CHIR99021 (Axon), 1 μM PD0325901 (Axon), and 1 : 5,000 hLIF. The culture plates were pre-treated with a 0.01% poly-L-ornithine solution (Sigma).

Plasmids

The plasmid pRosa26-GOF-2APuro-MUT was constructed based on the plasmid Rosa26-GOF-2APuro described earlier [17]. pRosa26-GOF-2APuro-MUT carries a 9.8-kb fragment of the *Pou5f1* gene, including its proximal and distal enhancers, homology arms targeting the *Rosa26* locus, and a gene coding for resistance to a selectable marker, puromycin. A point synonymous mutation was introduced into the PAM sequence of the first exon of *Pou5f1* within the plasmid pRosa26-GOF-2APuro to prevent knockout of exogenous *Pou5f1*.

The plasmid pRosa26-GR-Cdx2 carrying the Cdx2 sequence "fused" to the ligand-binding domain of the glucocorticoid receptor (GR) was ligated using constructs obtained earlier [18]. This plasmid also carries the gentamicin resistance gene and homology arms targeting the *Rosa26* locus. A sequence of guide RNA (gRNA) 5'-ACTCCAGTCTTTCTAGAAGA-3' paired with Cas9 nickase was used to incorporate the constructs into the alleles of the *Rosa26* locus.

CRISPR/Cas9-mediated *Pou5f1* knockout was performed using gRNA 5'- ACTCGTATGCGGGCGGACAT-3' encoded by the pX330-U6-Chimeric_BB-CBh-hSpCas9-EGFP vector. The gRNA sequences were selected using Benchling, an online platform (www.benchling.com).

Generating mutant ESC lines

In the first step of the generation of the $Pou5f1^{-/-}$;$Rosa26^{Pou5f1/Cdx2}$ ESC line, the $Pou5f1^{+/+}$;$Rosa26^{Pou5f1/+}$ line was used in order to produce cells with the *Pou5f1* sequence placed in the *Rosa26* locus and carrying a synonymous substitution within the first exon of *Pou5f1* (the pRosa26-GOF-2APuro-MUT vector being utilized as a donor sequence). Next, to perform an endogenous *Pou5f1* knockout, $Pou5f1^{+/+}$;$Rosa26^{Pou5f1/+}$ ESCs were transfected with the gRNA-/Cas9-encoding plasmid. Transfection was conducted using FuGene HD (Promega), in accordance with the manufacturer's protocol. The knockout of endogenous *Pou5f1* alleles and intact state of the exogenous construct within the *Rosa26* locus were verified by Sanger sequencing of TA-cloned alleles (*Fig. 2*) that involved cloning amplicons of these alleles into the pAL2-T vector (Evrogen).

In order to generate $Pou5f1^{-/-}$;$Rosa26^{Pou5f1/Cdx2}$ and $Pou5f1^{\Delta/\Delta}$;$Rosa26^{Pou5f1/Cdx2}$ ESCs, the GR-Cdx2 sequence was incorporated into the second *Rosa26* allele of the aforementioned ESC lines. The pRosa26-GR-Cdx2 vector was used as a donor sequence. Colonies were selected during six days using the geneticin antibiotic (G418) at a concentration of 500 μg/mL.

Trophoblast differentiation

The $Pou5f1^{-/-}$;$Rosa26^{Pou5f1/Cdx2}$ and $Pou5f1^{\Delta/\Delta}$;$Rosa26^{Pou5f1/Cdx2}$ ESC lines were cultured in the S/L medium supplemented with G418 (500 μg/mL, Neofroxx) and puromycin antibiotics (1 μg/mL, Sigma-Aldrich). The cells were reverted to their naïve state by culturing under 2i/L conditions for 7 days and then passaged into wells coated with an

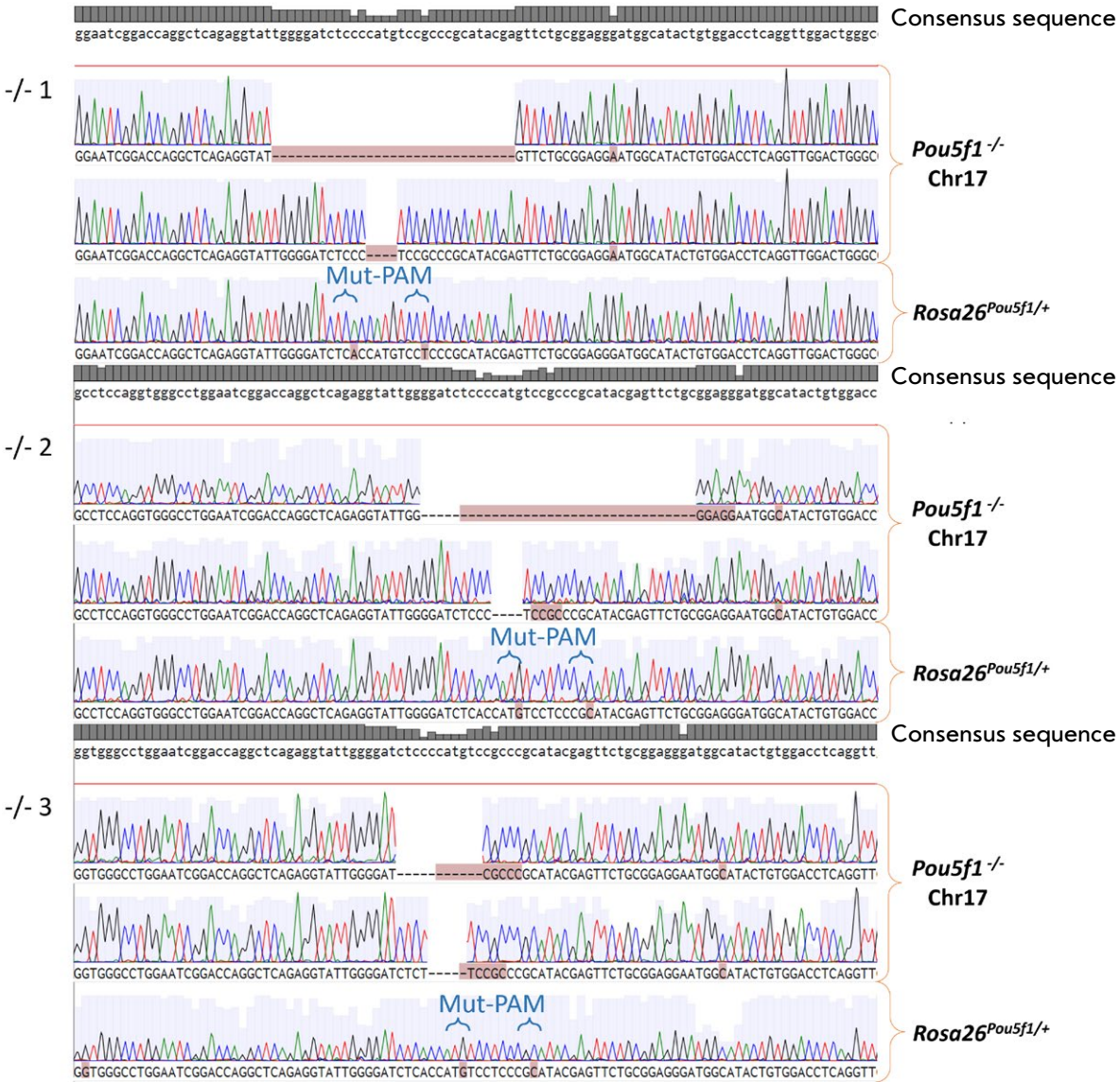


Fig. 2. Sequences of endogenous *Pou5f1* alleles from the *Pou5f1*$^{-/-}$;*Rosa26*$^{Pou5f1/+}$ cell line for three biological replicates. Note: "-/-" 1–3 – numbers of biological replicates for *Pou5f1*$^{-/-}$;*Rosa26*$^{Pou5f1/+}$ ESCs

MMC–MEF layer, then cultured in the TS medium based on a RPMI 1640 medium (Gibco) supplemented with 20% HyClone FBS (Cytiva), 1 mM sodium pyruvate (Gibco), 1× penicillin/streptomycin (Gibco), 0.1 mM β-mercaptoethanol (Sigma-Aldrich), 2 mM L-glutamine (Gibco), 1 μg/mL heparin (Hep) (Sigma-Aldrich), and 25 ng/mL fibroblast growth factor 4 (Fgf4) (Peprotech). The medium was pre-conditioned on MMC-MEFs for 72 h. A mixture of conditioned and fresh media at a 7 : 3 ratio was used for cell culturing. Dexamethasone (1 μM, Belmedpreparaty) and G418 (500 μg/mL, NeoFroxx) were added to the cells the next day after passaging. Four days later, the cells were reinoculated and cultured either in the standard TS medium or in the inflammation-mimicking TS medium. The latter was supplemented with either 300 U/mL interferon-gamma (IFNγ, ProSpec) or 1 μg/mL *E. coli* lipopolysaccharide (LPS, Sigma-Aldrich). Expression of trophoblast markers in the cells was analyzed one day after eliciting a pro-inflammatory response.

Quantitative RT-PCR

RNA was isolated using an RNA Solo kit (Evrogen); 1 μg of total RNA was utilized for cDNA synthesis. cDNA was synthesized in the presence of a

RiboCare RNase inhibitor and MMLV reverse transcriptase (Evrogen). Real-time PCR was conducted on a LightCycler® 96 system (Roche) using 5× qPCRmix-HS SYBR (Evrogen). Primer specificity and the optimal annealing temperatures (T_a) were pre-verified by PCR and electrophoresis using 4% agarose gel. *Table 1* lists the primer sequences and the selected T_a values. The *GAPDH* housekeeping gene was utilized as a reference gene. At least three biological replicates and two technical replicates were used for each cell line.

RESULTS

Generation of control *Pou5f1* knockout ESC lines

In order to investigate the *cis*-regulatory role of the *Pou5f1* promoter region in ESCs and their differentiated derivatives, we used the previously generated ESC line carrying a Cre-mediated deletion of the *loxP*-flanked promoter and the first exon of the *Pou5f1* gene. These cells maintain pluripotency owing to the expression of an exogenous *Pou5f1* fragment inserted into the *Rosa26* locus (*Pou5f1*$^{\Delta/\Delta}$;*Rosa26*$^{Pou5f1/+}$) [17]. The deletion in this cell line is identical to that introduced when studying the role of the transcription factor Oct4 in mouse cellular models of atherosclerosis (smooth muscle and endothelial cells) [15, 16]. We complemented this cell line with a new control line, *Pou5f1*$^{-/-}$;*Rosa26*$^{Pou5f1/+}$, where endogenous *Pou5f1* had been knocked out via indel mutations in the first exon. Like for the *Pou5f1*$^{\Delta/\Delta}$;*Rosa26*$^{Pou5f1/+}$ cell line, Oct4 expression was maintained via a 9.8-kb *Pou5f1* fragment inserted into one of the *Rosa26* alleles (*Fig. 2A*). This approach helped to eliminate the variability of Oct4 expression between the two ESC lines. This variability would inevitably arise when using the *Pou5f1*$^{\Delta/+}$ cell line. Importantly, the *Pou5f1*$^{-}$ allele had retained an intact promoter, making it possible to compare its functions directly with those of the *Pou5f1*$^{\Delta}$ allele. Previously, we have found that the *Rosa26*^{Pou5f1} allele can ensure self-maintenance of *Pou5f1*$^{\Delta/\Delta}$;*Rosa26*$^{Pou5f1/+}$ ESCs; however, these cells are unable to differentiate properly because the 9.8-kb *Pou5f1* fragment lacks all the essential *cis*-regulatory elements responsible for proper gene regulation during differentiation [17]. Therefore, directed differentiation of *Pou5f1*$^{\Delta/\Delta}$;*Rosa26*$^{Pou5f1/+}$ and *Pou5f1*$^{-/-}$;*Rosa26*$^{Pou5f1/+}$ ESCs represented a separate problem that needed to be addressed in this study.

Assessment of the ability of generated ESCs to differentiate into the trophoblast lineage

We chose the trophoblast lineage to differentiate ESCs into. It is known that trophoblast cells, which ultimately segregate at the late blastocyst stage as trophectoderm, endow maternal immune tolerance to the fetus after implantation by actively synthesizing non-classical MHC molecules [19]. Furthermore, trophoblast segregation is accompanied by *Pou5f1* silencing [20], which may trigger promoter activity switch from regulating *Pou5f1* itself to regulating the neighboring MHC-cluster genes [21]. Therefore, we concluded that trophoblast differentiation may serve as a suitable model for assessing gene expression profiles within the *Pou5f1-MHC* locus.

The differentiation protocol was based on forced expression of Cdx2, a key master regulator of trophoblast development [22, 23], which was also inserted into the *Rosa26* locus. The approach was chosen as the most straightforward alternative to those relying on media and growth factors, owing to its simplicity and the available published protocols. For controlled trophoblast differentiation, we used Cdx2 as a component of the fusion protein containing a ligand-binding domain of a glucocorticoid receptor (GR) that was activated by adding dexamethasone (Dex) to the medium. *Figure 3A* shows the final configurations of the *Pou5f1*$^{\Delta/\Delta}$;*Rosa26*$^{Pou5f1/Cdx2}$ and *Pou5f1*$^{-/-}$;*Rosa26*$^{Pou5f1/Cdx2}$ ESC lines.

Since the efficiency of trophoblast differentiation of ESCs under forced Cdx2 expression depends on the pluripotent stage [24], at the initial differentiation stage, *Pou5f1*$^{\Delta/\Delta}$;*Rosa26*$^{Pou5f1/Cdx2}$ and *Pou5f1*$^{-/-}$;*Rosa26*$^{Pou5f1/Cdx2}$ ESCs were reverted to their naïve state by 7-day culturing in the 2i/L medium. Furthermore, this experimental timepoint was used for monitoring changes in gene expression over time. The second and hinge study point was on Day 6 of cell culturing in the presence of dexamethasone, corresponding to Day 14 of the entire experiment (*Fig. 3B*).

By Day 6 of culturing in the presence of Dex, the cells, which originally had had a dome-shaped (under the SL conditions) or spherical (under the naïve 2iL conditions) colony shape, had morphed into flat colonies with clearly defined borders and an angular cell morphology, resembling those previously described for trophoblast stem cells [22, 23] (*Fig. 4A*).

An analysis of the marker expression profile on Day 6 of differentiation in the presence of Dex revealed a significant decline in the Oct4 mRNA level (compared to that in naïve ESCs) and an increase in the levels of trophectoderm marker mRNA in both cell lines. Mouse placenta was used as a control for the expression levels of trophoblast markers. The total Cdx2 levels in both ESC lines were significantly higher than that in the placenta. Differential analysis of endogenous Cdx2 and exogenous GR-Cdx2 mRNA levels established that this difference in the

Table 1. List of oligonucleotides used for quantitative real-time PCR

Primer	Nucleotide sequence 5'→3'	T, °C	Amplicon size, bp
qGAPDH-F	ACCCTTAAGAGGGATGCTGC	60	83
qGAPDH-R	CGGGACGAGGAAACACTCTC		
qOct4A-F	AGTGGAAAGCAACTCAGAGG	60	135
qOct4A-R	AACTGTTCTAGCTCCTTCTGC		
qCdx2-F	AGTCCCTAGGAAGCCAAGTGAA	60	96
qCdx2-R	AGTGAAACTCCTTCTCCAGCTC		
qCdx2GR-F	GCTGAAATCATCACCAATCAGATAC	60	134
qCdx2GR-R	CGCACGGAGCTAGGATACAT		
qCdx2endo-F	AGGCTGAGCCATGAGGAGTA	60	125
qCdx2endo-R	ctGAGGTCCATAATTCCACTCA		
qMash2-F	CGGGATCTGCACTCGAGGATT	65	86
qMash2-R	CCCCGTACCAGTCAAGGTGTG		
qTcfap2C-F	CGTCTCTCGTGGAAGGTGAAG	60	114
qTcfap2C-R	CCCCAAGATGTGGTCTCGTT		
qHand1-F	CCTACTTGATGGACGTGCTGG	60	129
qHand1-R	TTTCGGGCTGCTGAGGCAAC		
qElf5-F	CATTCGCTCGCAAGGTTACT	60	133
qElf5-R	GAGGCTTGTTCGGCTGTGA		
qH2-K1-F	TCCACTGTCTCCAACATGGC	60	113
qH2-K1-R	CCACCTGTGTTTCTCCTTCTCA		
qH2-Q6,8-F	CTGACCCTGATCGAGACCCG	60	112
qH2-Q6,8-R	TGTCCACGTAGCCGACGATAA		
qH2-Q7,9-F	GAGCTGTGGTGGCTTTTGTG	68	85
qH2-Q7,9-R	TGTCTTCATGCTGGAGCTGG		
qH2-Q10-F	ACATTGCTGATCTGCTGTGGC	60	120
qH2-Q10-R	GTCAGGTGTCTTCACACTGGAG		
qH2-Dmb1-F	ATGGCGCAAGTCTCATTCCT	68	95
qH2-Dmb1-R	TCTCCTTGGTTCCGGGTTCT		
qH2-Bl-F	ACCGGCTCCAACATGGTAAA	60	114
qH2-Bl-R	AGGAAGGATGGCTATTTTTCTGCT		
qH2-T23-F	ATAGATACCTACGGCTGGGAAATG	60	105
qH2-T23-R	AGCACCTCAGGGTGACTTCAT		
qTcf19-F	GATGATGAGGTCTCCCCAGG	60	107
qTcf19-R	TTTCCCTGTGGTCATTCCCC		
qPsors1C2-F	CTGTGTGCAGGAGGCATTTC	68	86
qPsors1C2-R	AGGGATCACCAGGGATTGGG		
Gm32362-F	GTCTGGAGAACCAAAGACAGCA	60	114
Gm32362-R	TTACAGCTTGGGATGCTCTTC		
Prrc2a-F	GAGATCCAGAAACCCGCTGTT	60	104
Prrc2a-F	TTCAGGCTTGGAAGGTTGGC		
Neu1-F	CCGGGATGTGACCTTCGAC	60	127
Neu1-R	CAGGGTCAGGTTCACTCGGA		
TNF-F	GTGCCTATGTCTCAGCCTCTT	60	117
TNF-R	AGGCCATTTGGGAACTTCTCATC		

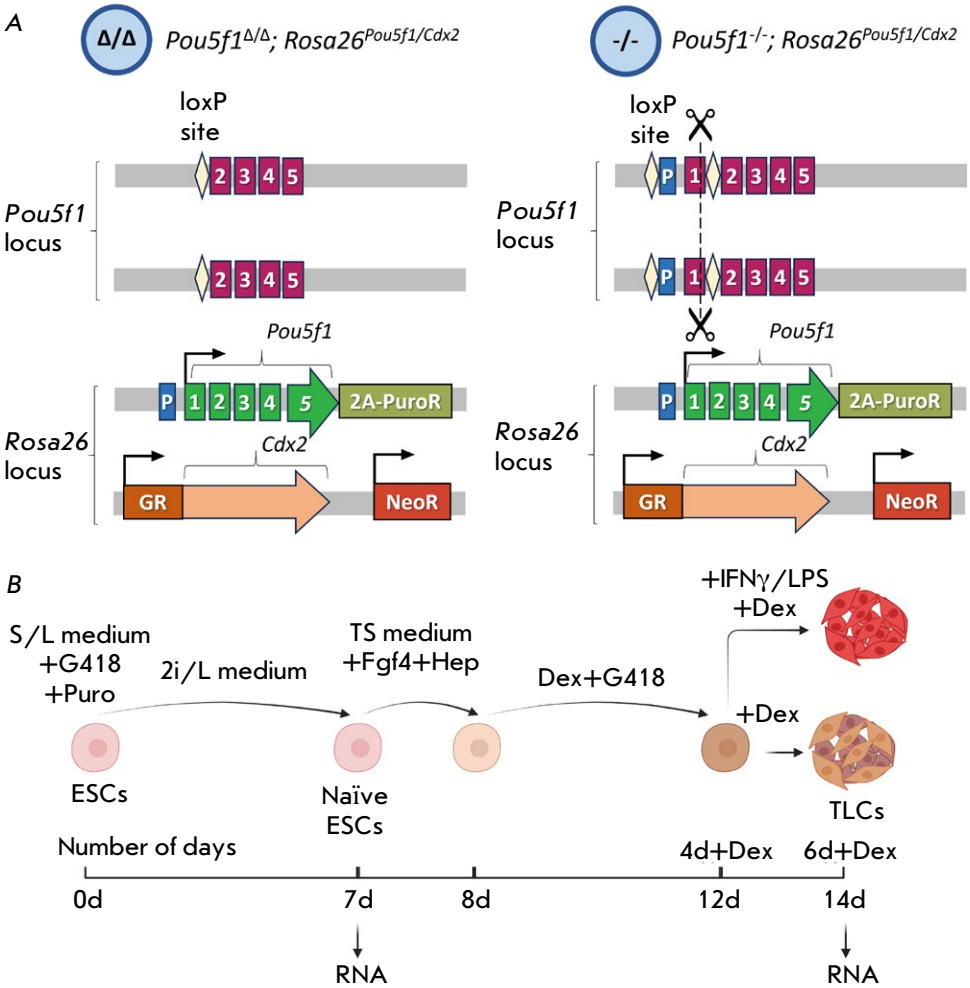


Fig. 3. Cell lines and the experimental protocol. (*A*) Schematic representation of the experimental embryonic stem cell (ESC) lines. "Δ/Δ" – *Pou5f1*$^{Δ/Δ}$; *Rosa26*$^{Pou5f1/Cdx2}$ ESC line with a deletion of the endogenous *Pou5f1* promoter; "-/-" – *Pou5f1*$^{-/-}$;*Rosa26*$^{Pou5f1/Cdx2}$ ESC line with an intact endogenous promoter and an inactivating indel mutation in the first exon of the gene. P – promoter; 1–5 – exons of the *Pou5f1* gene; 2A-PuroR – P2A site followed by the puromycin resistance gene PuroR; GR – ligand-binding domain of the glucocorticoid receptor; NeoR – the G418/neomycin resistance gene. (*B*) Schematic representation of ESC differentiation towards the trophoblast lineage (see the Materials and Methods section for a detailed description). Fgf4 – fibroblast growth factor 4; Hep – heparin; Dex – dexamethasone; IFNγ – interferon gamma; LPS – lipopolysaccharides; TLCs – trophoblast-like cells. The figure was created using BioRender

total Cdx2 levels was due to an induced overexpression of GR-Cdx2. Meanwhile, the endogenous Cdx2 level also increased to a level akin to that in placenta. We revealed no statistically significant differences in Cdx2 expression between the *Pou5f1*$^{Δ/Δ}$;*Rosa26*$^{Pou5f1/Cdx2}$ and *Pou5f1*$^{-/-}$;*Rosa26*$^{Pou5f1/Cdx2}$ ESCs, which is important for proper data interpretation. Moreover, expression of other trophectoderm markers (Tcfap2c, Mash2, and Hand1) was also demonstrated for the resulting trophoblast-like cells TLCs (*Fig. 4B*).

Assessment of the impact of the *Pou5f1* promoter region on gene expression within the *Pou5f1-MHC* locus

During the experiment, the cells were divided into groups and exposed to IFNγ or lipopolysaccharide (LPS). IFNγ and LPS are commonly utilized in various *in vitro* and *in vivo* inflammation models, so we addressed the hypothesis holding that induction of pro-inflammatory signals would promote the upregulation of the expression of immune-related genes, including the MHC genes, which would allow to more thoroughly assess the differences in the expression of the selected genes between generated cell lines. However, the differences in the expression of several MHC genes (*H2-K1, H2-T23, H2-Bl, H2-Dmb1, H2-Q6,8*, and *H2-Q7,9*) had been induced already by culture conditions, while their expression levels were identical in the *Pou5f1*$^{Δ/Δ}$;*Rosa26*$^{Pou5f1/Cdx2}$ and *Pou5f1*$^{-/-}$;*Rosa26*$^{Pou5f1/Cdx2}$ ESCs (*Fig. 5A*). *Tcf19* was the only gene whose expression was significantly different between the two genotypes (*Fig. 5B*). Notably, in undifferentiated *Pou5f1*$^{Δ/Δ}$;*Rosa26*$^{Pou5f1/Cdx2}$ ESCs cultured under naïve (2i/L) conditions, Tcf19 expression was already elevated compared to that of *Pou5f1*$^{-/-}$;*Rosa26*$^{Pou5f1/Cdx2}$ ESCs (*Fig. 5C*).

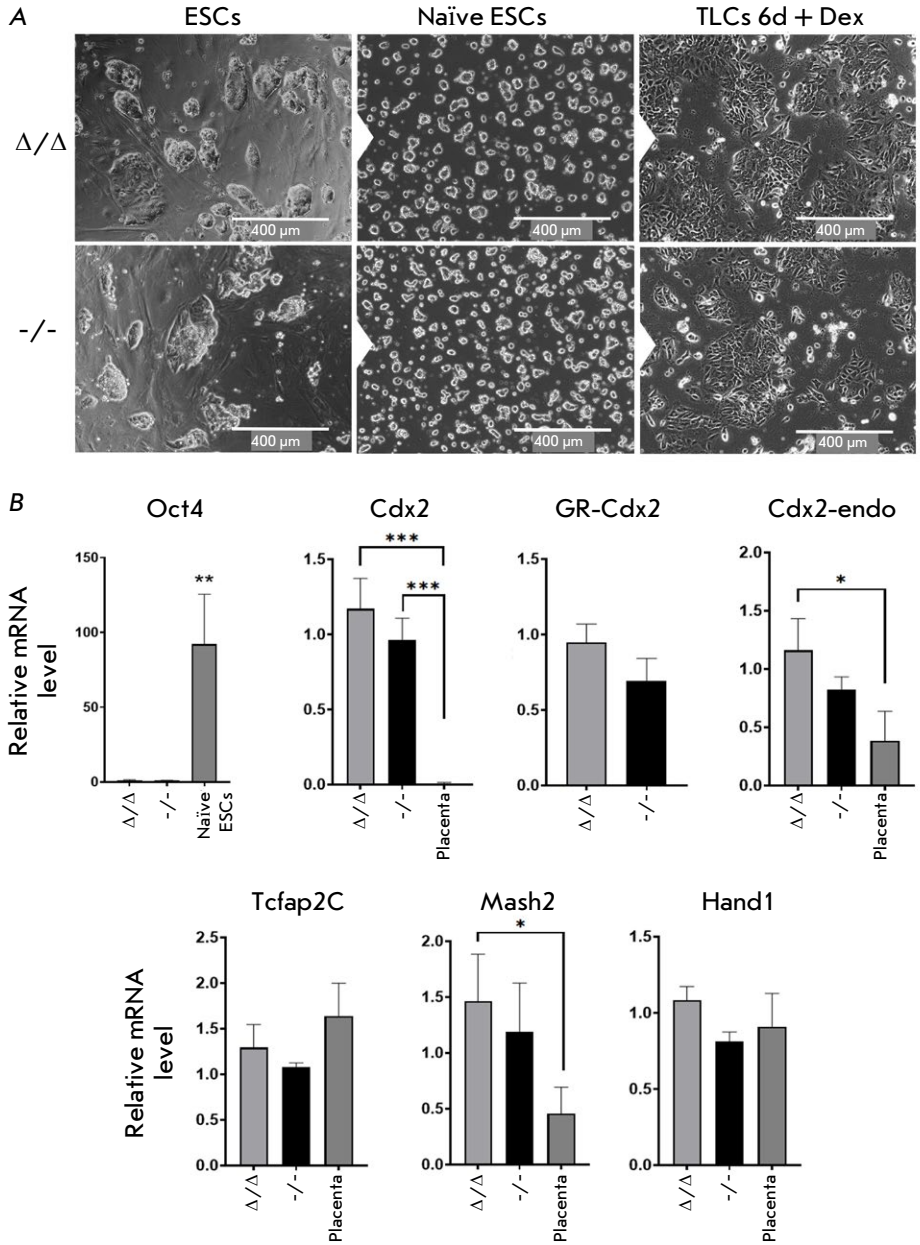


Fig. 4. Validation of the ability of *Pou5f1*$^{Δ/Δ}$;*Rosa26*$^{Pou5f1/Cdx2}$ and *Pou5f1*$^{-/-}$;*Rosa26*$^{Pou5f1/Cdx2}$ ESC lines to differentiate towards the trophoblast lineage. (*A*) Morphological characteristics of cells at different stages of differentiation: serum (S/L) culture conditions (left), naïve (2i/L) culture conditions (middle), and trophectoderm cells induced by Dex treatment for six days (right). (*B*) Analysis of the expression of trophoblast markers (Cdx2, Tcfap2C, Mash2, and Hand1) during differentiation compared to placenta. Designations are the same as those in *Fig. 3A*. $^{*}P \leq 0.05$; $^{**}P \leq 0.01$; $^{***}P \leq 0.001$ according to ANOVA

DISCUSSION

The question regarding the existence of *Pou5f1* expression outside the generally accepted concept of pluripotency remains to be addressed. The available evidence suggests that *Pou5f1* plays no functional role in differentiated mammalian cells, as indicated by the absence of phenotypic effects to the knockout of this gene and potential errors in the interpretation of the immunostaining and RT-PCR data [25–27]. On the other hand, recent research using functional genetic approaches convincingly demonstrates the role played by *Pou5f1* in somatic cells. Among those, there are studies describing the effect of *Pou5f1* knockout in smooth muscle and endothelial cells, as well as the study by Zalc et al., who had revealed *Pou5f1* reactivation in cranial neural crest cells and substantiated its role in enhancing the differentiation potential of these cells during embryogenesis [15, 16, 28].

Our hypothesis could integrate the reported findings from the perspective of the *cis*-regulatory properties of the *Pou5f1* promoter, confirming the activity of this gene on the one hand, while, on the other hand, decoupling it from the Oct4 protein, the product of this gene.

Elucidating the precise mechanism of how the *Pou5f1* gene functions in the context of atheroscle-

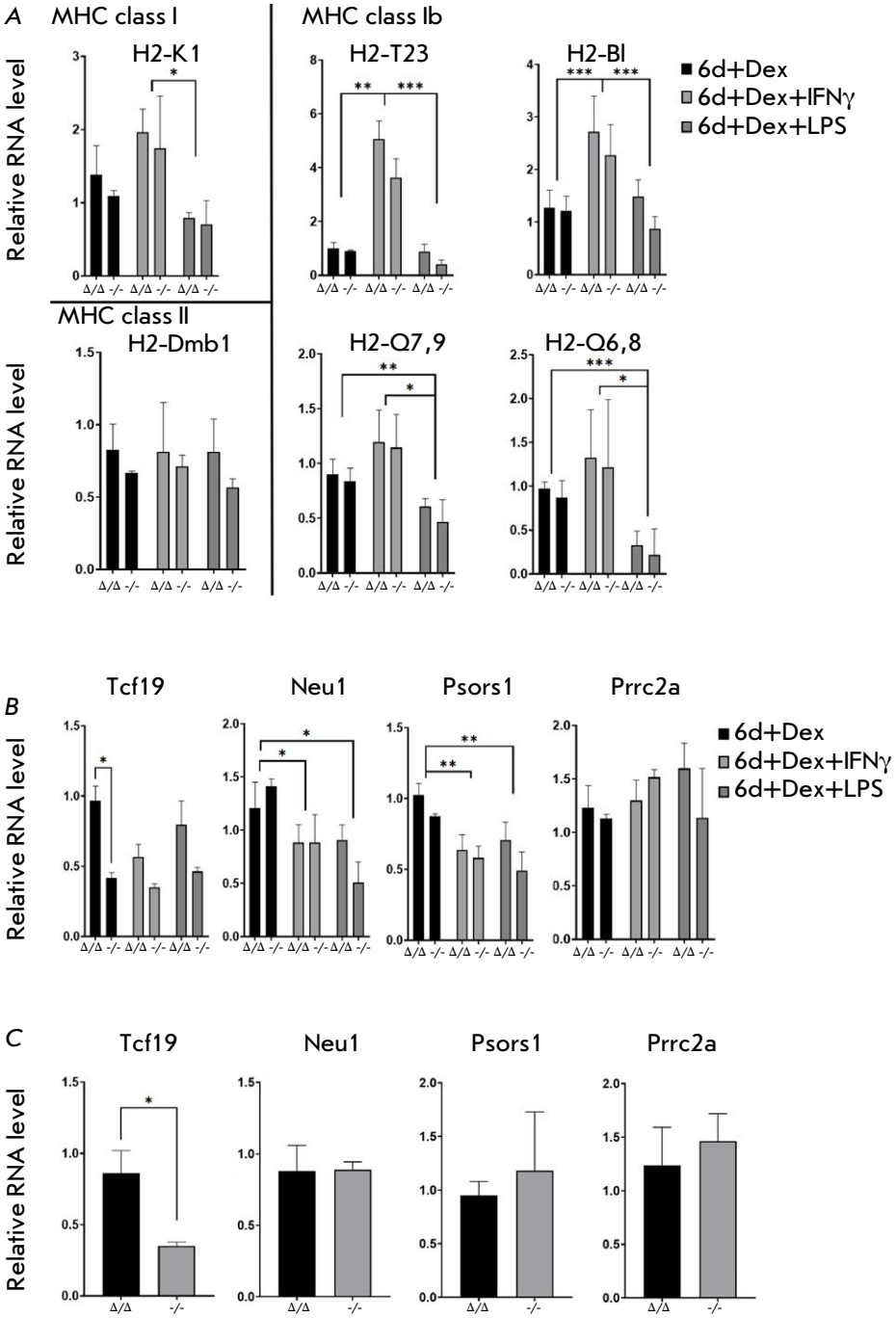


Fig. 5. Comparison of *Pou5f1-MHC* locus-related gene expression between the *Pou5f1*$^{\Delta/\Delta}$;*Rosa26*$^{Pou5f1/Cdx2}$ and *Pou5f1*$^{-/-}$;*Rosa26*$^{Pou5f1/Cdx2}$ cell lines under standard and pro-inflammatory culture conditions. (*A*, *B*) Comparison of the relative mRNA levels between the *Pou5f1*$^{\Delta/\Delta}$;*Rosa26*$^{Pou5f1/Cdx2}$ and *Pou5f1*$^{-/-}$;*Rosa26*$^{Pou5f1/Cdx2}$ ESC lines after six days of culture with dexamethasone (Dex) under standard and pro-inflammatory conditions (with IFNγ or LPS). Panel (*A*) presents the expression analysis of MHC class I and II genes; panel (*B*) compares the expression of the genes within the *Pou5f1-MHC* locus that were previously demonstrated to exhibit *cis*-regulatory activity towards *Pou5f1*. (*C*) Comparison of the expression of the genes from panel (*B*) in undifferentiated *Pou5f1*$^{\Delta/\Delta}$;*Rosa26*$^{Pou5f1/Cdx2}$ and *Pou5f1*$^{-/-}$;*Rosa26*$^{Pou5f1/Cdx2}$ ESCs cultured under 2i/L conditions. Figure legend is the same as that in *Fig. 3A*. $^{*}P \leq 0.05$; $^{**}P \leq 0.01$; $^{***}P \leq 0.001$ according to ANOVA. Comparisons were performed between the "Δ/Δ" and "-/-" cell lines under each culture condition, as well as between different conditions using the Tukey's test

rosis is a critical endeavor whose resolution is of certain importance not only for fundamental research, but also for potential medical applications. Thus, if the effects reported for atherosclerosis models have anything to do with the transcription factor Oct4, it should be regarded as a potential effector protein in the therapy of this disease. If the atherosclerotic phenotype is related to the *cis*-regulatory activity of the *Pou5f1* promoter, the focus of therapeutic strategies should be shifted toward the modulation of this activity.

Unlike the approach presented in this work, the earlier models for studying the *Pou5f1* gene were primarily designed to investigate its function in pluripotent stem cells, and the pluripotency of the cells was maintained using transgenic *Pou5f1* cDNA under the control of constitutive promoters [3, 29]. Not only did our approach allow us to generate an isogenic pair of

cell lines with *Pou5f1* expression inactivated during directed differentiation, but it also made it possible to compare them because of the identical location of exogenous *Pou5f1*, which would have been impossible if lentiviral vectors had been used. We believe that the developed model can help answer the question regarding *Pou5f1* expression in differentiated cells. The present study is the first step towards doing that. Although we did not observe any sweeping effect of *Pou5f1* promoter deletion on the expression of the genes within the *MHC* locus, one of the studied genes, *Tcf19*, was found to be susceptible to the introduced modifications. Interestingly, this gene is the nearest neighbor of *Pou5f1*, which may facilitate the interplay between their regulatory sequences. On the other hand, since the observed differences between the cell lines arise at the pluripotent stage, the mechanistic scenario for the effect of the introduced deletion can be considered definitely plausible. Thus, in the case of competition between the transcriptional machineries of the oppositely oriented *Tcf19* and *Pou5f1* genes, deletion of the *Pou5f1* promoter may relieve transcriptional interference, thereby favoring the expression of *Tcf19*. Although such a highly specific effect was unexpected, it appears to be consistent with the central concept of pluripotency. Being transcriptionally active in pluripotent cells, *Pou5f1* may, through alterations in its activity (e.g., due to specific mutations), affect the expression of *Tcf19*, potentially initiating a cascade of gene regulatory disruptions in daughter cells, including non-pluripotent ones. In turn, it may contribute to the development of various pathologies. This hypothesis offers a plausible explanation for the findings obtained in studies that have focused on *Pou5f1* polymorphisms associated with psoriasis [12], especially taking into account the association between *Tcf19* and this disease [30, 31]. Interestingly, *Tcf19* may also be involved in inflammatory responses, thus linking our findings to the data obtained using atherosclerosis models [32, 33]. A point of difference lies in the fact that *Pou5f1* knockout in those models was conditional; i.e., it was induced specifically in vascular smooth muscle or endothelial cells. Nonetheless, it remains possible that even the deletion of a methylated *Pou5f1* region could enhance *Tcf19* expression, which requires further investigation.

CONCLUSIONS

In this study, we developed a unique genetic model for investigating the role of the *Pou5f1* promoter sequence in the regulation of the expression of the genes that do not play a crucial role in pluripotent cells, providing a tool for uncovering potential non-classical functions of *Pou5f1* in differentiated cells. We have partially confirmed the hypothesis on the *cis*-regulatory activity of the *Pou5f1* promoter region with respect to the genes residing within the *Pou5f1-MHC* locus (to be more precise, with respect to its nearest neighbor, the *Tcf19* gene). Future research will focus on refining the regulatory landscape of the *Pou5f1-MHC* locus in other types of differentiated cells. ●

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-15-2021-1075 dated September 28, 2021) for obtaining and cultivation of cell lines, as well as by the Russian Science Foundation (grant No. 24-75-10131, https://rscf.ru/project/24-75-10131/) for differentiation and qRT-PCR.

REFERENCES

1. Takahashi K., Yamanaka S. // Cell. 2006. V. 126. № 4. P. 663–676. doi: 10.1016/j.cell.2006.07.024.
2. Niwa H., Miyazaki J., Smith A.G. // Nat. Genet. 2000. V. 24. № 4. P. 372–376. doi: 10.1038/74199.
3. Radzisheuskaya A., Le Bin Chia G., Dos Santos R.L., Theunissen T.W., Castro L.F.C., Nichols J., Silva J.C.R. // Nat. Cell Biol. 2013. V. 15. № 6. P. 579–590. doi: 10.1038/ncb2742.
4. Strebinger D., Deluz C., Friman E.T., Govindan S., Alber A.B., Suter D.M. // Mol. Syst. Biol. 2019. V. 15. № 9. P. 9002. doi: 10.15252/msb.20199002.
5. Horton R., Wilming L., Rand V., Lovering R.C., Bruford E.A., Khodiyar V.K., Lush M.J., Povey S., Conover C.J., Wright M.W., et al. // Nat. Rev. Genet. 2004. V. 5. № 12. P. 889–899. doi: 10.1038/nrg1489.
6. Shiina T., Inoko H., Kulski J. // Tissue Antigens. 2004. V. 64. № 6. P. 631–649. doi: 10.1111/j.1399-0039.2004.00327.x.
7. Nichols J., Smith A. // Cell Stem Cell. 2009. V. 4. № 6. P. 487–492. doi: 10.1016/j.stem.2009.05.015.
8. Choi H.W., Joo J.Y., Hong Y.J., Kim J.S., Song H., Lee J.W., Wu G., Schöler H.R., Do J.T. // Stem Cell Repts. 2016. V. 7. № 5. P. 911–926. doi: 10.1016/j.stemcr.2016.09.012.
9. Yeom Y.I., Fuhrmann G., Ovitt C.E., Brehm A., Ohbo K., Gross M., Hübner K., Schöler H.R. // Development. 1996. V. 122. № 3. P. 881–894. doi: 10.1242/dev.122.3.881.
10. Diao Y., Fang R., Li B., Meng Z., Yu J., Qiu Y., Lin K.C., Huang H., Liu T., Marina R.J., et al. // Nat. Methods. 2017. V. 14. № 6. P. 629–635. doi: 10.1038/nmeth.4264.
11. Canver M.C., Tripathi P., Bullen M.J., Olshansky M., Kumar Y., Wong L.H., Turner S.J., Lessard S., Pinello L., Orkin S.H., et al. // J. Biol. Chem. 2020. V. 295. № 47. P. 15797–15809. doi: 10.1074/jbc.RA120.013772.
12. Chang Y.T., Hsu C.Y., Chou C.T., Lin M.W., Shiao Y.M., Tsai C.Y., Yu C.W., Shiue J.J., Lee Y.F., Huang C.H., et al. // J. Dermatol. Sci. 2007. V. 46. № 2. P. 153–156. doi: 10.1016/j.jdermsci.2007.01.003.
13. Liu X., Li W., Fu X., Xu Y. // Front. Immunol. 2017. V. 8.

P. 645. doi: 10.3389/fimmu.2017.00645.
14. Drukker M., Katz G., Urbach A., Schuldiner M., Markel G., Itskovitz-Eldor J., Reubinoff B., Mandelboim O., Benvenisty N. // Proc. Natl. Acad. Sci. USA. 2002. V. 99. № 15. P. 9864–9869. doi: 10.1073/pnas.142298299.
15. Cherepanova O.A., Gomez D., Shankman L.S., Swiatlowska P., Williams J., Sarmento O.F., Alencar G.F., Hess D.L., Bevard M.H., Greene E.S., et al. // Nat. Med. 2016. V. 22. № 6. P. 657–665. doi: 10.1038/nm.4109.
16. Shin J., Tkachenko S., Chaklader M., Pletz C., Singh K., Bulut G.B., Han Y.M., Mitchell K., Baylis R.A., Kuzmin A.A., et al. // Cardiovasc. Res. 2022. V. 118. № 11. P. 2458–2477. doi: 10.1093/cvr/cvac036.
17. Kuzmin A.A., Ermakova V.V., Potapenko E.V., Ostroverkhova M.G., Guriev N.A. Tomilin A.N. // J. Dev. Biol. 2020. V. 51. № 6. P. 410–415. doi: 10.1134/S106236042006003X.
18. Tolkunova E., Cavaleri F., Eckardt S., Reinbold R., Christenson L.K., Schöler H.R., Tomilin A. // Stem Cells. 2006. V. 24. № 1. P. 139–144. doi: 10.1634/stemcells.2005-0240.
19. Rodgers J.R., Cook R.G. // Nat. Rev. Immunol. 2005. V. 5. № 6. P. 459–471. doi: 10.1038/nri1635.
20. Wu G., Schöler H.R. // Cell Regen. 2014. V. 3. № 1. P. 1–10. doi: 10.1186/2045-9769-3-7.
21. Malfait J., Wan J., Spicuglia S. // BioEssays. 2023. V. 45. № 10. P. 2300012. doi: 10.1002/bies.202300012.
22. Tanaka S., Kunath T., Hadjantonakis A.K., Nagy A., Rossant J. // Science. 1998. V. 282. № 5396. P. 2072–2075. doi: 10.1126/science.282.5396.2072.
23. Kehler J., Tolkunova E., Koschorz B., Pesce M., Gentile L., Boiani M., Lomelí H., Nagy A., McLaughlin K.J., Schöler H.R., et al. // EMBO Rep. 2004. V. 5. № 11. P. 1078–1083. doi: 10.1038/sj.embor.7400279.
24. Blij S., Parenti A., Tabatabai-Yazdi N., Ralston A. // Stem Cells Dev. 2015. V. 24. № 11. P. 1352–1365. doi: 10.1089/scd.2014.0395.
25. Lengner C.J., Camargo F.D., Hochedlinger K., Welstead G.G., Zaidi S., Gokhale S., Schöler H.R., Tomilin A., Jaenisch R. // Cell Stem Cell. 2007. V. 1. № 4. P. 403–415. doi: 10.1016/j.stem.2007.07.020.
26. Liedtke S., Enczmann J., Waclawczyk S., Wernet P., Kögler G. // Cell Stem Cell. 2007. V. 1. № 4. P. 364–366. doi: 10.1016/j.stem.2007.09.003.
27. Warthemann R., Eildermann K., Debowski K., Behr R. // Mol. Hum. Reprod. 2012. V. 18. № 12. P. 605–612. doi: 10.1093/molehr/gas032.
28. Zalc A., Sinha R., Gulati G.S., Wesche D.J., Daszczuk P., Swigut T., Weissman I.L., Wysocka J. // Science. 2021. V. 371. № 6529. P. eabb4776. doi: 10.1126/science.abb4776.
29. Karwacki-Neisius V., Göke J., Osorno R., Halbritter F., Ng J.H., Weiße A.Y., Wong F.C., Gagliardi A., Mullin N.P., Festuccia N., et al. // Cell Stem Cell. 2013. V. 12. № 5. P. 531–545. doi: 10.1016/j.stem.2013.04.023.
30. Nedoszytko B., Szczerkowska-Dobosz A., Stawczyk-Macieja M., Owczarczyk-Saczonek A., Reich A., Bartosińska J., Batycka-Baran A., Czajkowski R., Dobrucki I.T., Dobrucki L.W., et al. // Adv. Dermatol. Allergol. 2020. V. 37. № 3. P. 283–298. doi: 10.5114/ada.2020.96243.
31. Ling Y.H., Chen Y., Leung K.N., Chan K.M., Liu W.K. // PLoS One. 2023. V. 18. № 12. P. e0294661. doi: 10.1371/journal.pone.0294661.
32. Yang G.H., Fontaine D.A., Lodh S., Blumer J.T., Roopra A., Davis D.B. // Metabolites. 2021. V. 11. № 8. P. 513. doi: 10.3390/metabo11080513.
33. Ma X., Wang Q., Sun C., Agarwal I., Wu H., Chen J., Zhao C., Qi G., Teng Q., Yuan C., et al. // Cell Rep. 2023. V. 42. № 8. P. 112944. doi: 10.1016/j.celrep.2023.112944.

Classification and Quantification of Unproductive Splicing Events

L. G. Zavileyskiy*, E. A. Chernyavskaya, M. A. Vlasenok, D. D. Pervouchine
Center for Molecular and Cellular Biology, Skolkovo Institute of Science and Technology, Moscow, 121205 Russia
*E-mail: l.zavileisky@skoltech.ru
Received: November 26, 2024; in final form, March 31, 2025
DOI: 10.32607/actanaturae.27572

ABSTRACT **In eukaryotic cells, the nonsense-mediated decay (NMD) pathway degrades mRNAs with premature stop codons. The coupling between NMD and alternative splicing (AS) generates NMD-sensitive transcripts (NMD targets, NMDTs) that play an important role in the gene expression regulation via the unproductive splicing mechanism. Understanding this mechanism requires proper identification of NMDT-generating AS events. Here, we developed NMDj, a tool for the identification, classification and quantification of NMDT-generating AS events which does not rely on the best matching transcript partner principle employed by the existing methods. Instead, NMDj uses a set of characteristic introns that discriminate NMDTs from all protein-coding transcripts. The benchmark on simulated RNA-Seq data demonstrated that NMDj allows to quantify NMDT-generating AS events with better precision compared to other existing methods. NMDj represents a generic method suitable for the accurate classification of arbitrarily complex AS events that generate NMDTs. The NMDj pipeline is available through the repository https://github.com/zavilev/NMDj/.**
KEYWORDS **unproductive splicing, nonsense mediated decay, NMD, splicing, regulation.**
ABBREVIATIONS **NMD – Nonsense Mediated Decay; NMDT – NMD target transcript; PTC – premature termination codon; AS – alternative splicing; UTR – untranslated region; nt – nucleotide.**

INTRODUCTION

Eukaryotic cells express a large number of transcripts from each gene through alternative splicing (AS). By rough estimates, human protein-coding genes produce as many as ~150,000 expressed transcripts, an average of 7.4 isoforms per gene [1]. However, only half of these transcripts encode full-length proteins, while the remaining part may contain premature termination codons (PTC) [1, 2]. In eukaryotes, such transcripts are selectively eliminated by the pathway called the nonsense-mediated decay (NMD) [3].

In recent studies, it has been proposed that NMD not only prevents the translation of truncated proteins resulting from nonsense mutations and splicing errors, but is also involved in a wide range of biological processes, including gene expression regulation [4]. Most RNA-binding proteins (RBPs) control their own expression levels through a negative feedback loop in which the gene product binds to its cognate mRNA and induces AS that generates a PTC [5, 6]. It has been suggested that much of the impact of AS on the eukaryotic transcriptional landscape is mediated by the generation of NMD isoforms to limit gene expression, rather than the expansion of proteome diversity [2].

Local splicing changes, that is, the ones confined to a local region in the pre-mRNA, are one of the main sources of transcripts that are NMD targets (NMDT). Among the main types of local AS events, one can distinguish the so-called poison and essential exons which lead to the generation of NMDT upon exon inclusion and skipping, respectively, as well as the use of alternative 5'- and 3'-splice sites and intron retention [7]. Some of them (for instance, intron retention) may be involved in a particular biological process or may be preferentially regulated by the same splicing factor [8, 9]. However, the diversity of AS events is not limited to the main types listed above [6]. The task of characterizing complex AS events leading to the emergence of NMDT appears in many studies related to gene expression regulation [10–12].

To date, the only solution to this problem has been provided by the NMD Classifier [13]. Its approach is based on the assumption of minimal evolution/regulation, according to which NMDTs are the result of evolutionary or regulatory events that alter minimally the reading frame of a protein-coding transcript. That is, NMD Classifier finds the most similar coding transcript (in terms of shared nucleotide sequence) for each NMDT and considers the differences between

the best partner transcript and NMDT which cause a frameshift to be the generating AS event. However, the probability of NMDT being derived from a protein-coding transcript via AS depends not only on the similarity in their exon-intron architectures but also on their expression levels. The coding transcript with the highest expression level is more likely to be the source of NMDT [14]. Furthermore, NMDT may be derived from different transcripts with comparable expression levels, which calls into question the validity of the approach based on the selection of only one matching transcript partner.

In revisiting this problem, we developed NMDj, a tool for systematic search, classification and quantification of NMDT-generating AS events which takes into account all annotated transcripts and reports all introns that distinguish NMDTs from protein-coding transcripts. NMDj provides a more detailed classification of NMDT-generating AS events than the NMD Classifier. The coupling between NMD and AS is a crucial post-transcriptional mechanism of gene expression regulation [15]. Therefore, developing a method for searching, classifying, and quantifying AS events leading to NMDT which would take into account all the diversity of transcript isoforms is challenging. The NMDj method is aimed at tackling exactly this problem. It receives a set of transcripts in the form of an annotation database or transcript models constructed from RNA sequencing data as input, and provides the characterization of NMDT-generating AS events and their quantification as output.

EXPERIMENTAL

Genome annotation

The annotations of the human (GRCh38, version 108), mouse (mm10, version 113), zebrafish (danRer11, version 113), and Drosophila (dm6, version 113) genomes were downloaded from Ensembl in GTF format [16]. Only the transcripts of protein-coding genes with at least one annotated NMDT were considered. Transcripts without an annotated start or stop codon were filtered out. Genes without either NMDT or protein-coding transcripts were not considered.

NMD Classifier

The NMD Classifier source code was downloaded from [13]. To quantify local splicing alterations, the NMD Classifier output was converted to a list of alternative splice junctions corresponding to the four main types of AS events: alternative exons, alternative 5'- and 3'-splice sites, and intron retention (NMD_in, NMD_ex, A5SS, A3SS, NMD_IR, nNMD_IR).

The NMDj pipeline

The pipeline departs from a transcript annotation file in GFF/GTF format [17]. The following four features ("transcript", "exon", "start_codon" and "stop_codon") and three attributes ("gene_id", "transcript_id", "transcript_type") are considered. In addition to the main GFF/GTF file, NMDj can also accept a secondary input containing "transcript" and "exon" features, along with the "transcript_id" attribute. In this case, each transcript from the additional file is assigned to a gene from the main file based on the maximum number of common introns and a sequence overlap of at least 50%. For transcripts that were assigned to genes, the longest open reading frame is selected from those containing the annotated start codons and the corresponding start and stop codon positions are added to the annotation. As in Ensembl [18], a transcript is annotated as NMDT if there is an intron at least 50 nt downstream of the stop codon position.

Next, for each NMDT, NMDj considers the genomic interval spanning from the last splice site shared by NMDT and any protein-coding transcript with the same phase, or start codon in the absence of such, to the 3'end of the exon with PTC, or the closest downstream transcript end, if NMDT shares its stop codon with a protein-coding transcript. The characteristic introns are defined as all introns overlapping the genomic interval of interest except those shared by the NMDT and any coding transcript. The NMDT-generating AS event is defined as the set of characteristic introns described above. AS events from a pair of NMDTs are merged into a cluster if the NMDTs share at least one characteristic intron.

To classify NMD-generating AS events, NMDj by default uses MANE-Select transcripts as a reference, since they tend to be the most expressed ones [19]. However, a user-defined input can also be provided. NMDj builds a directed acyclic splicing graph using the splice sites of NMDT and splice sites of the reference transcript as nodes and introns and exons as edges, and it searches for "bubbles" defined by vertex-independent paths that contain characteristic introns [20, 21]. NMDj reports all found pairs of vertex-independent paths in the following form: $X_1...X_n : Y_1...Y_m$, where X_i and Y_j are "D" (donor) and "A" (acceptor) symbols, and $X_i \neq X_j$ and $Y_i \neq Y_j$ when $j = i \pm 1$. If the reference transcript set has not been specified, then NMDj iteratively compares the NMDT with each protein-coding transcript.

The last, optional step is the quantification of AS events using RNA-seq split read counts (the input table must be provided). NMDj computes the Ψ (percent-spliced-in) values, which estimate the expression

level of the NMDT relative to all the transcripts of the gene. It is calculated using the formula

$$\Psi = \frac{\sum_{i=1}^{A} a_i k_i}{\sum_{i=1}^{A} a_i k_i + \sum_{j=1}^{B} b_j r_j},$$

where A and B are the number of characteristic introns supporting NMDT and protein-coding transcripts, respectively; a_i and b_j are the number of RNA-seq split-reads aligned to the respective introns, and k_i and r_j are the weights that account for the number of times the characteristic introns occur in NMDT and coding transcripts, respectively. The weights k_i and r_j are computed independently for NMDT and coding transcripts. The natural requirement that the sum of the weights of the characteristic introns of each transcript be equal to 1 leads us to a system of n linear equations with m unknowns, where n is the number of transcripts and m is the number of characteristic introns. By the construction of characteristic introns, such a system is always consistent, but it can also have an infinite number of solutions. In general, one could make an unambiguous choice of k_i and r_j by imposing regularization constraints on this system. However, in NMDj we use the following heuristic algorithm, which allows us to define the value of Ψ in accordance with the existing definitions for the main types of AS events [6, 11].

Transcripts annotated in the interval are represented as a graph with the vertices being characteristic introns, and the edges being the exons (or their groups) that connect them. This graph is searched for pairs of vertices connected by only vertex-independent paths. For each such path, the weights of characteristic introns are assumed to be equal to each other. For a poison exon, for instance, there will be two such paths: one corresponding to exon inclusion (with two characteristic introns, each with a weight of 0.5), and another corresponding to exon skipping (with one characteristic intron, the weight of which is equal to 1). After the coefficients of the nodes between the identified pair are assigned values, these nodes are merged into one and the search in the new graph continues. At each step, the coefficients of the characteristic introns combined into a node are multiplied by the value assigned to that node and the procedure continues until all nodes are merged into one. This algorithm works for all simple types of AS events, and for complex AS events it works only under the assumption that all vertex-independent paths are nested.

The real and simulated RNA-seq data

To realistically model RNA sequencing data using known transcript expression levels, and hence relative NMDT expression levels, we selected three random samples in each of the three tissues (Muscle, Liver, and Cerebellum) using the panel of transcriptomic data from the Genotype-Tissue Expression project (GTEx [22]. The choice of the tissues was motivated by the fact that they differ most drastically in terms of AS [23, 24]. Transcript expression levels in the selected samples were obtained by rsem-calculate-expression with the --estimate-rspd option [25]. The expression levels of NMDTs, best partner transcripts, and MANE-Select transcripts as a fraction of the total gene expression were calculated for each gene. Sampling was repeated five times, and the results were averaged.

RNA-seq data simulation was performed by rsem-simulate-reads based on the transcript expression levels described above. For each sample, 50 mln paired-end reads were simulated. The simulated reads were aligned to the GRCh38 human genome using STAR aligner 2.7.3a [26]. Counts of split-reads were obtained using the IPSA package with default settings [27]. Transcript expression levels in the simulated samples were quantified by RSEM (as above) [25]; Salmon 1.10.3, with the options --seqBias --gcBias --posBias [28]; and StringTie 2.2.3, with the option -e [29]. To convert transcript-level quantification results to Ψ values of the AS events, the NMDT expression levels (in TPM, transcripts per million) were divided by the sum of expression levels of transcripts spanning the genomic regions found by NMDj.

RNA-seq data on NMD inactivation

The results of the experiments on the inactivation of NMD components (double knockdown of SMG6 and SMG7) followed by RNA-seq were obtained from Gene Expression Omnibus under the accession number GSE86148 in the FASTQ format and aligned to the human genome assembly GRCh38 (hg38) using the STAR aligner v2.7.8a in the paired-end mode. Counts of split-reads were obtained using the IPSA package in the default settings [27].

RESULTS

The NMDj pipeline

The NMDj pipeline consists of three main and three auxiliary steps (*Fig. 1A*). Starting from the transcript annotation database, it performs the reading frame search and predicts NMDT, if they are not annotated. NMDT are annotated based on the so-called 50-nt rule, which postulates that a transcript is recognized

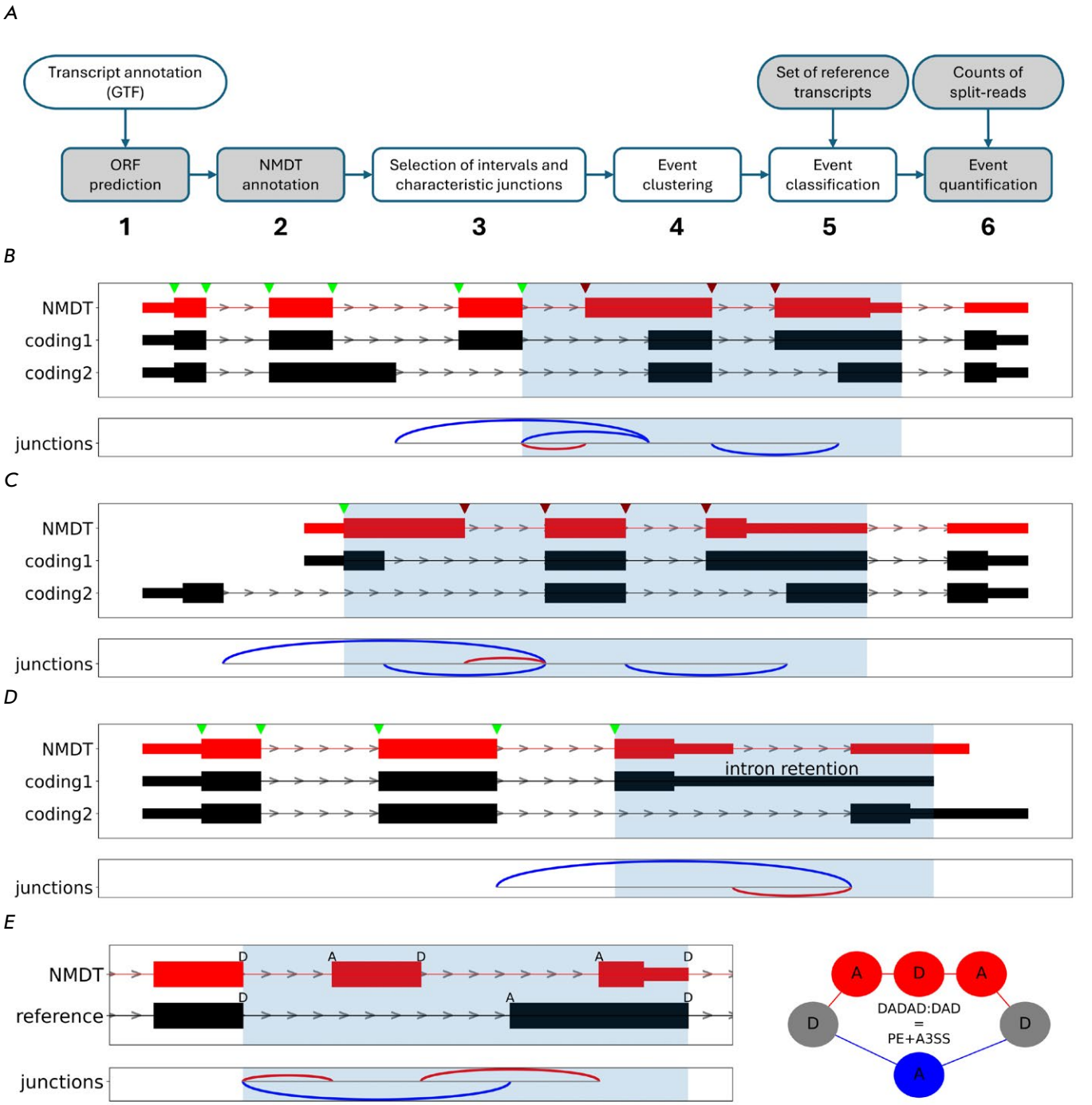


Fig. 1. The NMDj pipeline. (*A*) The pipeline flowchart. (*B*–*D*) The choice of interval boundaries (light blue shading). The 5′-boundary is either the last splice site common to NMDT and any coding transcript with the same phase (*B*), or the start codon if there is no such splice site (*C*). The 3′-boundary is either the donor splice site of the intron following the PTC-containing exon (B, C), or the end of the shortest 3′-UTR downstream of the NMDT stop codon (*D*). (*E*) An example of classification based on vertex-independent paths. NMDT and its reference coding transcript (left) correspond to a pair of vertex-independent paths consisting of donor and acceptor splice sites (right). NMDT and protein-coding transcripts, as well as their corresponding characteristic introns (arcs), are shown in red and blue, respectively. Splice sites of NMDT are indicated by green arrows if the NMDT frame matches the protein-coding frame, or red otherwise

Table 1. A list of NMDj event types and their synonyms in a classification provided by NMD Classifier

Type	NMDj	Description	Synonym
DADA:DA	PE	Poison cassette exon which triggers NMD upon inclusion	NMD_in
D(AD)nA:DA	PEn	*n* consecutive cassette exons which trigger NMD upon simultaneous inclusion	multi_NMD_in
DA:DADA	EE	Essential cassette exon which triggers NMD upon skipping	NMD_ex
DA:D(AD)nA	EEn	*n* consecutive cassette exons which trigger NMD upon simultaneous skipping	multi_NMD_ex
ADA:ADA	A5SS	Alternative 5'-splice sites	A5SS
DAD:DAD	A3SS	Alternative 3'-splice sites	A3SS
ADAD:ADAD	A5SS+A3SS	Both 5'- and 3'-splice sites of the same intron are alternative	A5SS,A3SS
AD:ADAD	IR	Intron retention which triggers NMD	nNMD_IR
ADAD:AD	ID	Intron excision which triggers NMD	NMD_IR
DADA:DADA	MXE	A pair of mutually exclusive adjacent exons	-
AD(AD)nA:ADA	A5SS+PEn	Alternative 5'-splice site and n consecutive poison exons	-
ADA:AD(AD)nA	A5SS+EEn	Alternative 5'-splice site and *n* consecutive essential exons	-
D(AD)nAD:DAD	PEn+A3SS	*n* consecutive poison exons and alternative 3'-splice site	-
DAD:D(AD)nAD	EEn+A3SS	*n* consecutive essential exons and alternative 3'-splice site	-
ADAD:AD(AD)nAD	A5SS+EE+A3SS	Alternative 5'-splice site, *n* consecutive essential exons and alternative 3'-splice site	-

as an NMD target if it contains an intron at least 50–55 nt downstream of the stop codon [30]. This rule departs from the assumption that exon junction complexes that are deposited on pre-mRNA during splicing are displaced during the pioneer round of translation, and ones that remain bound outside of the reading frame serve as a PTC signal [30]. In NMDj, we used the threshold of 50 nucleotides because this is the accepted value for automatic NMDT annotation in Ensembl [16]. However, the number of predicted NMDTs changes insignificantly when the threshold is increased to 55 nt (*Fig. S1*).

Once open reading frames are detected and NMDTs are annotated for each gene, NMDj begins searching for NMD-generating AS events. There exist multiple formalisms for describing AS events including binary classes (such as poison exons [31]), classification of connected components in a splice graph [32], and local splicing variations [33]. In this work, we define an AS event as a set of characteristic introns spanning the following genomic interval. For each NMDT, the 5'-boundary of the interval is defined to be the 3'-most splice site, which it shares with any protein-coding transcript with the same phase (*Fig. 1B*). If no such splice site exists, the 5' boundary is placed at the start codon of NMDT, if it is shared with at least one protein-coding transcript (*Fig. 1C*). The 3'-boundary of the interval is defined to be the 3'-end of the PTC-containing exon or, if NMDT shares its stop-codon with a protein-coding transcript, and in case it is not a true PTC, it is placed at the nearest transcript end (*Fig. 1D*).

Next, NMDj selects the characteristic introns that distinguish NMDT from protein-coding transcripts. All the introns that are adjacent to the interval or overlap with it, except the ones that are shared by the NMDT and at least one protein-coding transcript, are considered to be characteristic introns. As a result, each NMDT is characterized by a set of characteristic introns that originate either from it or from protein-coding transcripts (*Fig. 1B,D*, red and blue arcs). The characteristic introns are merged into clusters to reduce redundancy, as several NMDTs would often possess the same or very similar sets of characteristic introns.

NMDj classifies splicing events into major types such as poison (PE) and essential (EE) exons, alternative splice sites (A5SS, A3SS), and others (*Table 1*). The classification of the AS events is based on the concept of vertex-independent paths applied to splicing graphs [20, 34]. In a directed acyclic graph, whose nodes are donor (D) and acceptor (A) splice sites, and edges are exons and introns, one can define a vertex-independent path as a pair of paths that do not share

Table 2. NMD-generating AS events in the human and model organism's transcriptomes

	#Tr	#NMDT	NMDT, %	Fraction of AS events, %					
				PE	EE	A5SS	A3SS	IR	Other
Human	79940	16741	21	18	11	6	8	2	55
Mouse	49951	5339	11	18	18	11	14	4	36
Zebrafish	35040	854	2	11	10	11	12	23	32
Drosophila	30688	1325	4	18	4	12	9	16	41

Note: #Tr – total number of transcripts; #NMDT – number of NMDT; NMDT – fractions of NMTD (in %). Fractions (in %) of toxic (PE) and necessary (EE) exons, fractions of alternative 5'-(A5SS) and 3'-splicing sites (A3SS), fractions of retained introns (IR) and other events (Other).

any nodes except the first and last node (*Fig. 1E*). Each such pair is reported in a symbolic form representing the sequence of nodes; i.e., a poison exon (PE) corresponds to DADA:DA; an alternative 5'- splice-site (A5SS), to ADA:ADA; and multiple poison exons (PEn), to D(AD)nA:DA, where n is the number of exons. In the final step, NMDj quantifies each group of NMDTs by Ψ values based on split read counts from RNA-Seq experiments (see EXPERIMENTAL).

NMDj in application to human and model organism transcripts

The application of NMDj to annotated transcripts from human, mouse, zebrafish, and Drosophila showed that the proportion of NMDTs obeying the 50-nucleotide rule is significantly higher in humans and mice than it is in zebrafish and Drosophila, which is undoubtedly a result of differences in the quality and completeness of transcriptome annotations (*Table 2*). However, the frequencies of NMDT-generating AS events vary significantly between organisms. While in humans and mice NMDTs are generated more frequently through the use of poison and essential exons than they are through intron retention, in Drosophila and zebrafish the pattern is opposite. According to existing estimates, the proportion of intron retention among the major AS types is equally low in mammals as it is in other vertebrates and invertebrates [35]. Thus, the observed difference between NMDT-generating AS event frequencies can be explained neither by the different levels of abundance of their types nor by the different levels of completeness of the transcriptome annotation. Rather, the difference indicates the peculiarities of the NMD system's functioning in different taxonomic groups.

The advantages of NMDj in finding NMD-generating AS events

The existing approach to the analysis of NMD-generating AS events, which is implemented in the NMD Classifier, is based on choosing the best partner transcript. The main problem in this approach is that other transcripts and their expression levels are not taken into account when selecting the best partner transcript. A protein-coding transcript is unlikely to be the main source of NMDT if its expression level is low. To illustrate the importance of this issue, we applied the NMD Classifier to the Ensembl transcriptome annotation [16] and compared the identified set of best partner transcripts with those from the MANE-Select annotation considered as the set of the most expressed transcripts in each human gene [19].

MANE-Select transcripts were identified as best partners only for 25% of NMDTs, while they had a significantly higher expression level, as confirmed by a random sample of RNA-seq experiments from GTEx (*Fig. 2A*). Furthermore, when the best partner transcript was not MANE-Select, its contribution to the total gene expression level was comparable to that of NMDT. This suggests that the transcript that is most similar to NMDT in terms of the shared sequence can be at the same time a poor candidate for generating NMDT. Moreover, the MANE-Select transcripts are not always the most expressed ones. Tissues may differ in their most expressed transcripts (*Fig. 2B*) or express several transcripts at comparable levels. To address this, NMDj considers all annotated transcripts in order to avoid the problem of choosing one best transcript partner and clusters NMDTs with similar characteristic introns to obtain a concise and non-redundant set of AS events (*Fig. 2D*).

NMDj is particularly useful in genes with a complex splicing architecture. A notable example is *HPS1*, which contains a group of exons with lengths that are not multiples of three (*Fig. 2C*). Skipping of each single exon generates a NMDT, unless it is compensated by a downstream AS event that restores the coding frame. Simultaneous inclusion of exons 6a and 7 generates a NMDT. NMD Classifier selects the transcript with exon 5 as the best partner. This exon is

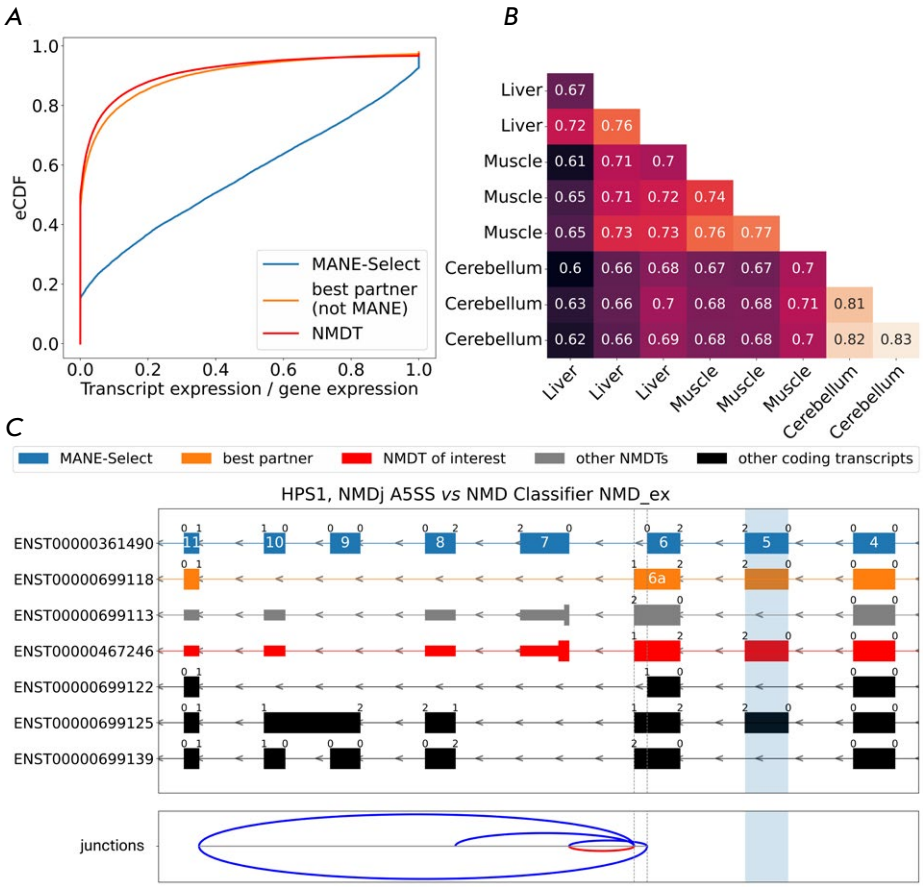


Fig. 2. NMDj and NMD Classifier best partner transcripts. (*A*) Transcript relative abundance (eCDF is the cumulative distribution function) estimated from a random sample of RNA-seq experiments from GTEx. (*B*) The proportion of genes whose most expressed transcripts match between pairs of GTEx tissue samples. (*C*) An example of a local NMD-generating event in the *HPS1* gene. The characteristic introns originating from NMDTs and protein-coding transcripts are shown by red and blue arcs, respectively. The phase of the reading frame is indicated above the exon boundaries. Transcript colors: MANE-Select (blue), NMDT (red), best partner transcript from NMD Classifier (orange), other transcripts (gray – NMDT, black — protein-coding). The essential exon predicted by the NMD classifier is highlighted in light-blue; however, NMDT is actually generated by a splice site shift in the MANE-Select isoform

skipped in the NMDT, which indeed disrupts the coding frame. However, it is also skipped in a protein-coding transcript, in which its frameshift is compensated by using exon 6 instead of exon 6a and skipping exons 7–10. NMDj correctly identifies the last splice site, in which the reading frame of NMDT matches that of a coding transcript, to be the 3'-boundary of exon 6a, which enables the detection of the only true NMD-generating AS event; namely, the splice junction between exons 6a and 7. It also identifies all alternative introns whose excision helps to bypass frame shifts. Interestingly, another NMDT with exon 5 included shares a characteristic intron with the previous one and is therefore clustered with it by NMDj.

NMDj provides a more detailed AS event classification

We compared the classification of AS events produced by NMDj and NMD Classifier in application to the same human transcriptome annotation (*Fig. 3A*). NMDj was configured to use MANE-Select transcripts as a reference. Out of 15,914 NMDTs, NMD Classifier and NMDj were able to classify AS events for 15,446 and 15,265 NMDTs, respectively. However, AS events were classified into the same type (*Table 1*) for only 60% of NMDTs.

While NMD Classifier subdivides AS events into a fixed number of most common types, NMDj is able to describe more complex splicing patterns using vertex-independent paths. In the *POR* gene, for example, NMDT differs from protein-coding isoforms by alternative 5'- and 3'-splice sites and a cassette exon (*Fig. 3B*). Such events tend to evade many standard tools for splicing analysis [31, 32]. The presence of AS types, which NMD Classifier is unable to properly detect, accounts for a large portion of inconsistencies between the two classifications. For example, a number of events classified by NMD Classifier as poison exons (NMD_in) are classified as PE+A3SS and MXE by NMDj (*Fig. 3A,C*). Another advantage of NMDj is the ability to classify AS events in 3'-untranslated regions (3'-UTRs). Among the events that induce NMD in the 3'-UTRs, the majority are expectedly represented by intron retention. Moreover, many 3'-UTR events do not intersect with the MANE-Select isoform (*Fig. 3A, S2*).

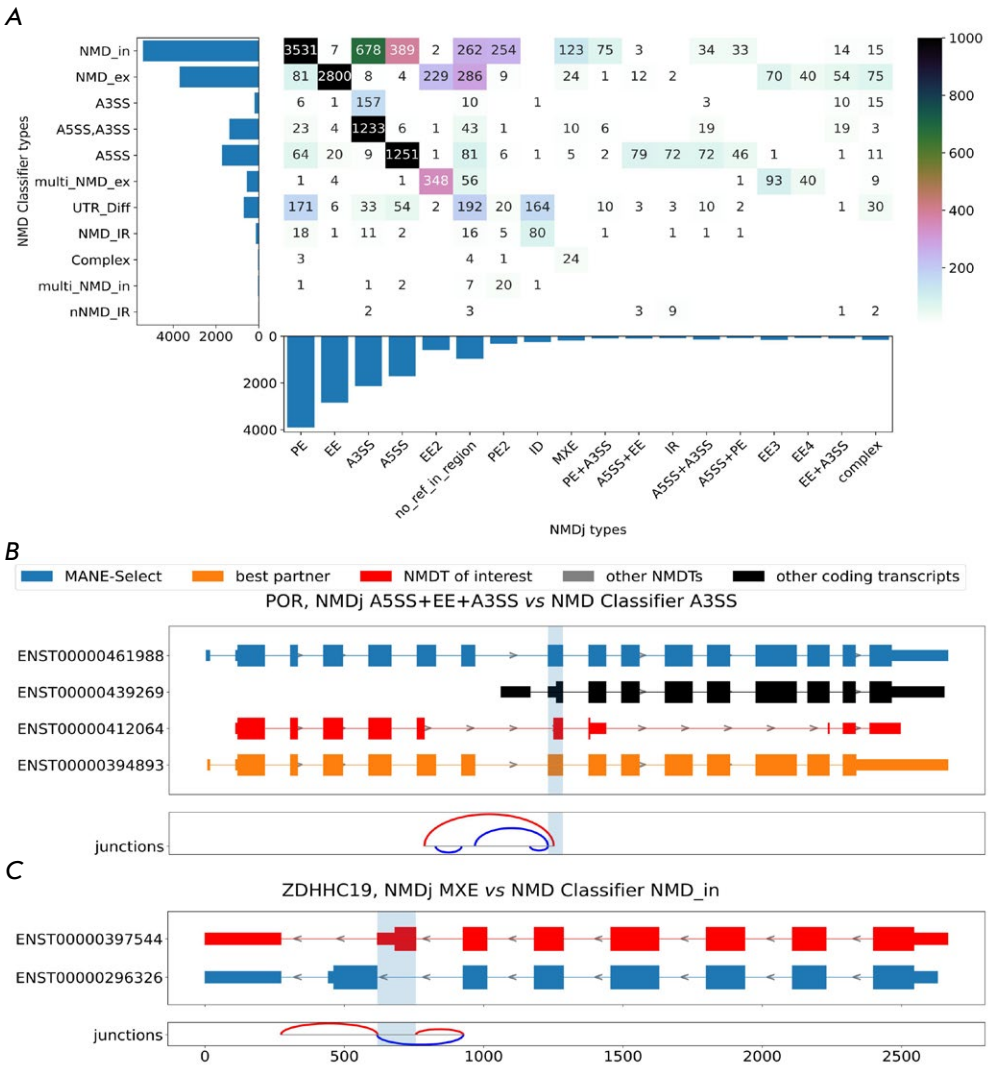


Fig. 3. AS event categorization. (*A*) A comparison of classifications by NMDj and NMD Classifier. Each cell represents the number of NMDTs classified into corresponding types by NMDj (rows) and NMD Classifier (columns). (*B*, *C*) Examples of rare NMD-generating AS events. The rest of the legend is as in *Fig. 2*

A relatively small number of other inconsistencies may be explained by the fact that NMDj and NMD Classifier use different reference transcripts to classify AS events. Only 61% of AS were classified in the same type when NMDj was configured to use best partner transcripts as a reference. A substantial portion of the differences seems to be the result of misclassification by NMD Classifier. For example, most events attributed to the "A3SS, A5SS" type by NMD classifier are classified as A3SS by NMDj (*Fig. 3A*). Meanwhile, the size of the NMD classifier's "A5SS,A3SS" class is far larger than the size of the "A3SS" class. This is counterintuitive, since the choice between a pair of alternative 5'-splice sites seems to be independent from the choice between a downstream pair of 3'-splice sites separated by a long intron [36]. Visual inspection of randomly selected individual cases of classification discrepancy confirmed the correctness of the classification provided by NMDj (*Fig. S3*).

NMDj benchmark on simulated and real data

NMD-generating AS events can be used to assess relative NMDT expression levels quantitatively using RNA-seq data. To evaluate the accuracy of NMDj in quantifying AS, we simulated RNA-seq reads based on the average transcript expression levels in human tissues. The estimated Ψ values computed from split reads aligned to characteristic introns were compared to the ground truth Ψ values, defined as the NMDT isoform abundance as a fraction of the total abundance of all transcripts of the given gene. As a measure of distance, we used the mean squared error (MSE) over all Ψ values across all the NMDT isoforms tested. It turned out that NMDj performed comparably to existing state-of-the-art methods for transcript-level quantification, while the MSE values for NMD classifier were substantially larger (*Fig. 4A*). Since the methods used to calculate the Ψ metric in NMDj and NMD Classifier were identical, this again

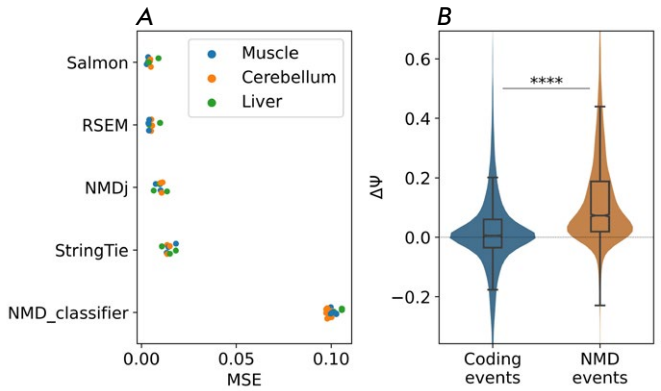


Fig. 4. Comparison of NMDj and NMD Classifier predictions. (*A*) Mean squared error (MSE) of Ψ values estimated by different methods from simulated RNA-Seq data relative to the ground truth values. (*B*) Statistically significant splicing changes of NMDT-generating and non-NMDT protein-coding AS events (cassette exons, alternative splice sites and intron retention) upon NMD inactivation by cycloheximide, quantified by NMDj. **** denotes statistically significant differences at the 0.1% significance level (Mann–Whitney test)

suggests that not only the best partner transcript but also other transcripts contribute significantly to the Ψ value.

To confirm that AS events predicted by NMDj actually generate NMDT, we compared the changes in the Ψ values of AS events generating and not generating NMDTs in NMD inactivation experiments implementing knockdown of its two key factors: SMG6 and SMG7 [14]. AS events that did not generate NMDT included cassette exons, alternative splice sites, and retained introns that had been found in non-NMDT protein-coding transcripts. As expected, upon inactivation of NMD, the Ψ values of NMD-generating AS events increased significantly more than did the Ψ values in coding transcripts (*Fig. 4B*).

DISCUSSION

The approach implemented in NMDj does not rely on a single best partner transcript, and that allows it to identify and properly describe many more NMD-generating AS events as compared to NMD Classifier. However, NMDj was unable to locate characteristic introns for some NMDTs (1,139 transcripts), which

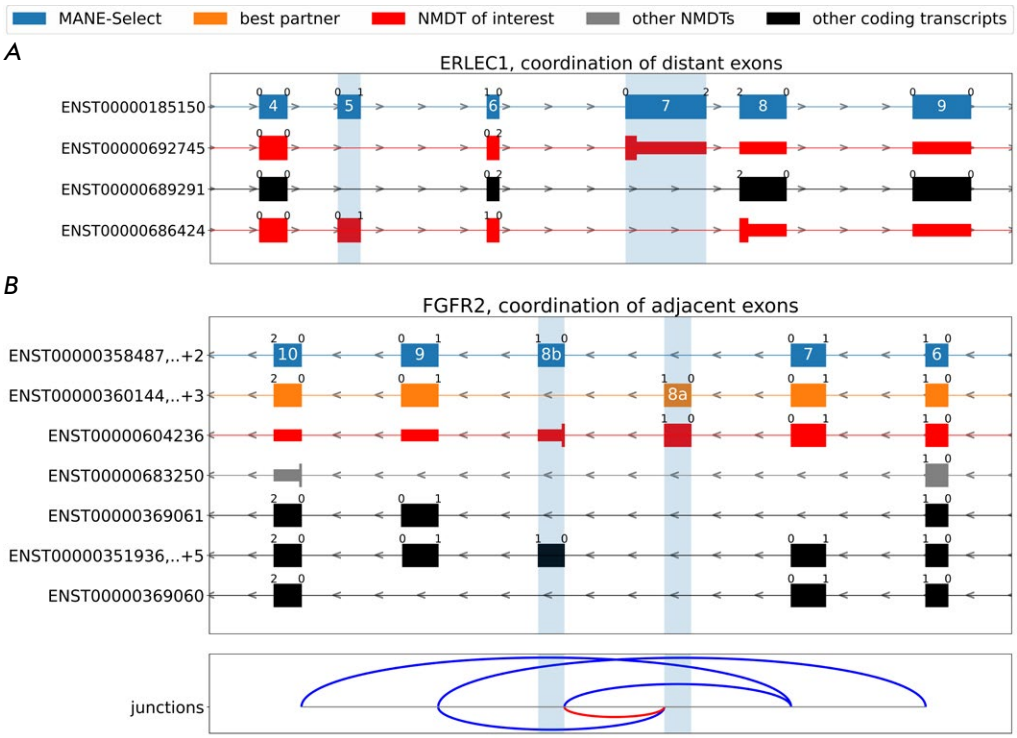


Fig. 5. (*A*) Coordinated splicing of distant cassette exons in the *ERLEC1* gene. (*B*) Coordinated splicing of adjacent exons in the *FGFR2* gene. Besides mutually exclusive splicing of exons 8a and 8b, transcript isoforms with coordinated skipping of exons 7–9 (NMDT) and ones with coordinated skipping of exons 7–8a,b and 8a,b–9 (protein-coding) are annotated. Simultaneous inclusion of exons 8a and 8b generates NMDT. Legend colors are as in *Fig. 2*

in most cases was the result of coordination between distant AS events and the usage of alternative start and stop codons. For instance, in the *ERLEC1* gene, simultaneous inclusion and exclusion of non-adjacent exons 5 and 7 preserves the reading frame, while inclusion of only one exon from the pair leads to NMDT (*Fig. 5A*). This example demonstrates that it is not always possible to establish a causal link between a particular local AS event and NMDT, because NMD sensitivity is a global property of a transcript which depends on coordination between distant AS events, while local AS events individually may not be capturing these global properties. Like other approaches that take into account only local AS events, NMDj is fundamentally incapable of correctly characterizing the cause of such NMDTs.

Local AS events are known to regulate gene expression by AS switching to NMDT production [5, 6]. Such switching is mediated by RNA-binding proteins that bind to the *cis*-elements in pre-mRNA and is typically regulated locally [37]. In contrast, little is known about the functional outcomes and exact regulatory mechanisms of coordination for AS events at large distances [38–41]. While the coordination between distant AS events could be important for producing protein isoforms with distinct functions, in some cases cells could use it to generate NMDTs. An example of this is the coordinated, mutually exclusive splicing of exons 8a and 8b in the *FGFR2* gene, which leads to functional protein products with different ligand specificities [41] (*Fig. 5B*). The inclusion of exon 8a is promoted by the epithelial-specific proteins ESRP1 and ESRP2, which bind to the same regulatory sequence inside the intron [42], but simultaneous inclusion of both exons generates NMDT. Thus, switching between FGFR2 isoforms is regulated on the level of local AS, while coordination of mutually exclusive exon choices is achieved by the elimination of an NMDT.

In sum, a simultaneous analysis of all splice isoforms, instead of single best-matching transcript partners, allows NMDj to identify NMD-generating AS events with higher accuracy. However, the technique shares a common limitation with other methods in classifying the coordinated action of distant AS events. Their analysis requires fundamentally different approaches. However, it seems more likely that NMD induces a non-random association of AS events than a regulated association of AS events induces NMD. Thus, the analysis of coordinated AS events falls outside the scope of this study for both technical and conceptual reasons.

The method developed in this paper can be used to study gene expression regulation via unproductive splicing [6]. In particular, it can be applied to problems such as searching for specifically expressed NMDTs and assessing the activity of the NMD system as a whole. Thus, NMDj closes the existing gap in the toolkit for studying the conjugation between AS and NMD. ●

The results presented here include data obtained from the GTEx Portal (dbGaP accession number phs000424/GRU).

This work was supported by the Russian Science Foundation grant (22-14-00330).

Supplementaries are available online at https://10.32607/actanaturae.27572.

REFERENCES

1. Tung K.-F., Pan C.-Y., Chen C.-H., Lin W.-C. // Sci. Rep. 2020. V. 10. № 1. P. 16245. doi: 10.1038/s41598-020-73081-5.
2. Fair B., Najar C.F.B.A., Zhao J., Lozano S., Reilly A., Mossian G., Staley J.P., Wang J., Li Y.I. // Nat. Genet. 2024. V. 56. № 9. P. 1851–1861. doi: 10.1038/s41588-024-01872-x.
3. Kurosaki T., Popp M.W., Maquat L.E. // Nat. Rev. Mol. Cell Biol. 2019. V. 20. № 7. P. 406–420. doi: 10.1038/s41580-019-0126-2.
4. Nasif S., Contu L., Mühlemann O. // Semin. Cell Dev. Biol. 2018. V. 75. P. 78–87. doi: 10.1016/j.semcdb.2017.08.053.
5. Müller-McNicoll M., Rossbach O., Hui J., Medenbach J. // J. Mol. Cell Biol. 2019. V. 11. № 10. P. 930–939. doi: 10.1093/jmcb/mjz043.
6. Zavileyskiy L.G., Pervouchine D.D. // Acta Naturae. 2024. V. 16. № 1. P. 4–13. doi: 10.32607/actanaturae.27337.
7. Yan Q., Weyn-Vanhentenryck S.M., Wu J., Sloan S.A., Zhang Y., Chen K., Wu J.Q., Barres B.A., Zhang C. // Proc. Natl. Acad. Sci. USA. 2015. V. 112. № 11. P. 3445–3450. doi: 10.1073/pnas.1502849112.
8. Wong J.J.-L., Ritchie W., Ebner O.A., Selbach M., Wong J.W.H., Huang Y., Gao D., Pinello N., Gonzalez M., Baidya K., et al. // Cell. 2013. V. 154. № 3. P. 583–595. doi: 10.1016/j.cell.2013.06.052.
9. Jangi M., Boutz P.L., Paul P., Sharp P.A. // Genes Dev. 2014. V. 28. № 6. P. 637–651. doi: 10.1101/gad.235770.113.
10. Margasyuk S., Kuznetsova A., Zavileyskiy L., Vlasenok M., Skvortsov D., Pervouchine D.D. // NAR Genom. Bioinform. 2024. V. 6. № 4. P. lqae163. doi: 10.1093/nargab/lqae163.
11. Mironov A., Petrova M., Margasyuk S., Vlasenok M., Mironov A.A., Skvortsov D., Pervouchine D.D. // Nucleic Acids Res. 2023. V. 51. № 7. P. 3055–3066. doi: 10.1093/nar/gkad161.
12. Pervouchine D., Popov Y., Berry A., Borsari B., Frankish A., Guigó R. // Nucleic Acids Res. 2019. V. 47. № 10. P. 5293–5306. doi: 10.1093/nar/gkz193.

13. Hsu M.-K., Lin H.-Y., Chen F.-C. // PLoS One. 2017. V. 12. № 4. P. e0174798. doi: 10.1371/journal.pone.0174798.
14. Karousis E.D., Gypas F., Zavolan M., Mühlemann O. // Genome Biol. 2021. V. 22. № 1. P. 223. doi: 10.1186/s13059-021-02439-3.
15. Isken O., Maquat L.E. // Nat. Rev. Genet. 2008. V. 9. № 9. P. 699–712. doi: 10.1038/nrg2402.
16. Harrison P.W., Amode M.R., Austine-Orimoloye O., Azov A.G., Barba M., Barnes I., Becker A., Bennett R., Berry A., Bhai J., et al. // Nucleic Acids Res. 2024. V. 52. № D1. P. D891–D899. doi: 10.1093/nar/gkad1049.
17. Cunningham F., Allen J.E., Allen J., Alvarez-Jarreta J., Amode M.R., Armean I.M., Austine-Orimoloye O., Azov A.G., Barnes I., Bennett R., et al. // Nucleic Acids Res. 2022. V. 50. № D1. P. D988–D995. doi: 10.1093/nar/gkab1049.
18. Britto-Borges T., Gehring N.H., Boehm V., Dieterich C. // RNA. 2024. V. 30. № 10. P. 1277–1291. doi: 10.1261/rna.080066.124.
19. Morales J., Pujar S., Loveland J.E., Astashyn A., Bennett R., Berry A., Cox E., Davidson C., Ermolaeva O., Farrell C.M., et al. // Nature. 2022. V. 604. № 7905. P. 310–315. doi: 10.1038/s41586-022-04558-8.
20. Ivanov T.M., Pervouchine D.D. // Genes. 2018. V. 9. № 7. doi: 10.3390/genes9070356.
21. Ma C., Zheng H., Kingsford C. // Algorithms Mol. Biol. 2021. V. 16. № 1. P. 5. doi: 10.1186/s13015-021-00184-7.
22. Lonsdale J., Thomas J., Salvatore M., Phillips R., Lo E., Shad S., Hasz R., Walters G., Garcia F., Young N., et al. // Nat. Genet. 2013. V. 45. № 6. P. 580–585. doi: 10.1038/ng.2653.
23. Yeo G., Holste D., Kreiman G., Burge C.B. // Genome Biol. 2004. V. 5. № 10. P. R74.
24. Barbosa-Morais N.L., Irimia M., Pan Q., Xiong H.Y., Gueroussov S., Lee L.J., Slobodeniuc V., Kutter C., Watt S., Colak R., et al. // Science. 2012. V. 338. № 6114. P. 1587–1593. doi: 10.1126/science.1230612.
25. Li B., Dewey C.N. // BMC Bioinformatics. 2011. V. 12. P. 323. doi: 10.1186/1471-2105-12-323.
26. Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T.R. // Bioinformatics. 2013. V. 29. № 1. P. 15–21. doi: 10.1093/bioinformatics/bts635.
27. Pervouchine D.D., Knowles D.G., Guigó R. // Bioinformatics. 2013. V. 29. № 2. P. 273–274. doi: 10.1093/bioinformatics/bts678.
28. Patro R., Duggal G., Love M.I., Irizarry R.A., Kingsford C. // Nat. Methods. 2017. V. 14. № 4. P. 417–419. doi: 10.1038/nmeth.4197.
29. Pertea M., Pertea G.M., Antonescu C.M., Chang T.-C., Mendell J.T., Salzberg S.L. // Nat. Biotechnol. 2015. V. 33. № 3. P. 290–295. doi: 10.1038/nbt.3122.
30. Popp M.W., Maquat L.E. // Cell. 2016. V. 165. № 6. P. 1319–1322. doi: 10.1016/j.cell.2016.05.053.
31. Shen S., Park J.W., Lu Z., Lin L., Henry M.D., Wu Y.N., Zhou Q., Xing Y. // Proc. Natl. Acad. Sci. USA. 2014. V. 111. № 51. P. E5593–5601. doi: 10.1073/pnas.1419161111.
32. Li Y.I., Knowles D.A., Humphrey J., Barbeira A.N., Dickinson S.P., Im H.K., Pritchard J.K. // Nat. Genet. 2018. V. 50. № 1. P. 151–158. doi: 10.1038/s41588-017-0004-9.
33. Vaquero-Garcia J., Barrera A., Gazzara M.R., González-Vallinas J., Lahens N.F., Hogenesch J.B., Lynch K.W., Barash Y. // eLife. 2016. V. 5. P. e11752. doi: 10.7554/eLife.11752.
34. Wang X., Dalkic E., Wu M., Chan C. // Curr. Opin. Biotechnol. 2008. V. 19. № 5. P. 482–491. doi: 10.1016/j.copbio.2008.07.011.
35. Kim E., Magen A., Ast G. // Nucleic Acids Res. 2007. V. 35. № 1. P. 125–131.
36. Conti L.D., Baralle M., Buratti E. // Wiley Interdiscip. Rev. RNA. 2013. V. 4. № 1. P. 49–60. doi: 10.1002/wrna.1140.
37. Barash Y., Calarco J.A., Gao W., Pan Q., Wang X., Shai O., Blencowe B.J., Frey B.J. // Nature. 2010. V. 465. № 7294. P. 53–59. doi: 10.1038/nature09000.
38. Joglekar A., Foord C., Jarroux J., Pollard S., Tilgner H.U. // Transcription. 2023. V. 14. № 3-5. P. 92–104. doi: 10.1080/21541264.2023.2213514.
39. Fededa J.P., Petrillo E., Gelfand M.S., Neverov A.D., Kadener S., Nogués G., Pelisch F., Baralle F.E., Muro A.F., Kornblihtt A.R. // Mol. Cell. 2005. V. 19. № 3. P. 393–404.
40. Bushra S., Lin Y.-N., Joudaki A., Ito M., Ohkawara B., Ohno K., Masuda A. // Int. J. Mol. Sci. 2023. V. 24. № 8. P. 7420. doi: 10.3390/ijms24087420.
41. Holzmann K., Grunt T., Heinzle C., Sampl S., Steinhoff H., Reichmann N., Kleiter M., Hauck M., Marian B. // J. Nucleic Acids. 2012. V. 2012. P. 950508. doi: 10.1155/2012/950508.
42. Warzecha C.C., Sato T.K., Nabet B., Hogenesch J.B., Carstens R.P. // Mol. Cell. 2009. V. 33. № 5. P. 591–601. doi: 10.1016/j.molcel.2009.01.025.

The Hypomethylating Agent 5-Azacitidine Potentiates the Effect of RAS and Sp1 Inhibitors in Neuroblastoma Cells

K. A. Ivanenko[1*], A. V. Snezhkina[1], M. A. Zolotovskaia[2,3], P. V. Spirin[1,4], O. G. Leonova[1], V. I. Popenko[1], A. V. Kudryavtseva[1,4], A. A. Buzdin[2,3,5,6], V. S. Prassolov[1,4], T. D. Lebedev[1,4]

[1]Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia
[2]Moscow Institute of Physics and Technology, Dolgoprudny, 141701 Russia
[3]Sechenov First Moscow State Medical University, Moscow, 119991 Russia
[4]Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia
[5]Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Moscow, 117997 Russia
[6]PathoBiology Group, European Organization for Research and Treatment of Cancer (EORTC), Brussels, 1200 Belgium
[*]E-mail: karina.ivanenko@mail.ru

Received November 08, 2024; in final form, April 03, 2025
DOI: 10.32607/actanaturae.27558

ABSTRACT **Neuroblastoma is a malignant solid tumor caused by the transformation of neural crest cells. Neuroblastoma predominantly occurs in children and is associated with a poor prognosis. In this regard, the development of novel approaches to neuroblastoma treatment, including combination therapy, is relevant. DNA hypermethylation of neuroblastoma cells indicates that it is possible to use hypomethylating agents in a combination therapy of the disease. In order to identify effective combinations of antitumor drugs, we analyzed the transcriptomic changes that take place in neuroblastoma SH-SY5Y cells after treatment with the hypomethylating agent 5-azacitidine and then experimentally tested the effectiveness of these combinations. Mithramycin A and lonafarnib were the two drugs that, in combination with 5-azacitidine, appeared to exert a synergistic effect on SH-SY5Y cell death. These drugs inhibit the signaling pathway associated with the transcription factor Sp1 and RAS-MAPK signaling pathway, which are activated by 5-azacitidine. An analysis of the signaling pathways also revealed an activation of the signaling pathways associated with neuroblastoma cell differentiation, as well as apoptosis induction, as confirmed by multiplex and confocal microscopy. Hence, by analyzing the changes in the signaling pathways, the mechanisms of cell death and cell adaptation to hypomethylating agents can be understood, and this can be further used to develop novel therapeutic approaches to neuroblastoma therapy.**

KEYWORDS **pediatric malignant diseases, combination therapy, epigenetic regulators.**

ABBREVIATIONS **5-Aza – 5-azacitidine; GD2 – disialoganglioside; ALK – anaplastic lymphoma kinase; MDM2 – murine double minute 2; DMSO – dimethyl sulfoxide; PKB – protein kinase B; NGF – nerve growth factor; ILK – integrin-linked kinase; TRK – tropomyosin receptor kinase; IGF1R – insulin-like growth factor 1; MAPK – mitogen-activated protein kinase; ERK – extracellular signal-regulated kinase; EGFR – epidermal growth factor receptor; FGFR – fibroblast growth factor receptor; JAK – Janus kinase; CHK – checkpoint kinase; mTOR – mammalian target of rapamycin; RAF – Rapidly Accelerated Fibrosarcoma, serine/threonine protein kinase; CDK – cyclin-dependent kinase; RTK – receptor tyrosine kinase; RAS – Rat Sarcoma, small G protein.**

INTRODUCTION

Neuroblastoma is an extracranial solid tumor that is the result of malignant transformation of neural crest cells during the formation of the sympathetic nervous system [1]. The five-year survival rate for children with high-risk neuroblastoma (50% of the cases) is approximately 60% [2]. The main treatment modalities for high-risk neuroblastoma include intensive chemotherapy, radiation therapy, autologous stem cell transplantation, and immunotherapy [3]. Targeted agents are under development: they would target disialoganglioside (GD2) [4], anaplastic lymphoma kinase (ALK) [5–7], E3 ubiquitin-protein ligase (MDM2) [8], and components of the signaling pathways such as the PI3K/Akt/mTOR, Fos/Jun, and RAS-MAPK pathways [9]. ALK inhibitors are already undergoing clinical trials for the treatment of patients with recurrent and refractory neuroblastoma [10]. The existing treatment approaches to such patients sometimes fail the test of effectiveness; therefore, combination therapies for neuroblastoma are now being pursued [11].

Alterations in DNA methylation are frequently observed in malignant cells of different origins, as well as hypermethylation of tumor suppressor promoters or global hypomethylation, in particular [12]. Two DNA methyltransferase inhibitors, 5-azacitidine (5-Aza) and its analog decitabine, have been approved for the treatment of myelodysplastic syndromes [13, 14]. 5-Aza is a hypomethylating agent and a synthetic analog of cytidine. Incorporating 5-Aza into DNA disrupts the activity of DNA methyltransferases, resulting in DNA hypomethylation and damage. The drug has been approved for the treatment of patients with acute myeloid leukemia and myelodysplastic syndromes [15].

Genomic DNA hypermethylation in neuroblastoma cells is associated with a poor prognosis [16]. 5-Aza was shown to induce the differentiation of neuroblastoma cells, reduce proliferation and colony formation, and to potentiate the cytotoxic effects of agents such as doxorubicin, cisplatin, and etoposide [17]. Decitabine has previously been tested in combination with doxorubicin; however, phase I clinical trials revealed the high toxicity associated with this combination [18]. Inhibitors of epigenetic regulators, such as histone deacetylase inhibitors, may exhibit synergism when used in combination with receptor tyrosine kinase inhibitors by upregulating their expression [19]. Additionally, 5-Aza can significantly affect the expression of the genes involved in oncogenesis through DNA demethylation. Therefore, it appears opportune to explore new therapeutic approaches that are based on the combination of 5-Aza with other antitumor agents.

This study analyzed the changes in gene expression and the activity of the signaling pathways in neuroblastoma SH-SY5Y cells exposed to 5-Aza in order to identify the most effective combinations of 5-Aza with various antitumor agents. The functional significance of alterations in the signaling pathway activity at the transcriptomic level was additionally examined by investigating intracellular processes using fluorescence microscopy and assessing the synergistic effects of 5-Aza and inhibitors of different signaling pathways. These findings can be used as a platform for developing novel therapeutic approaches to treat neuroblastomas susceptible to demethylating agents.

MATERIALS AND METHODS

Cell cultures and inhibitors

Cell lines derived from human malignant tumors, including neuroblastoma SH-SY5Y, breast cancer SK-BR-3, renal cell carcinoma 786-O, cervical cancer SiHa, and ovarian cancer SK-OV-3 cells, were cultured in RPMI-1640 medium (Capricorn Scientific, Germany). Colorectal carcinoma HCT-116, lung adenocarcinoma H1299, glioblastoma LN-18, and rhabdomyosarcoma TE-671 cells were cultured in DMEM medium (Capricorn Scientific). All the cell lines were cultured at 37°C in a humidified atmosphere containing 5% CO_2, supplemented with 10% fetal bovine serum (Gibco, USA), 1 mM sodium pyruvate (Gibco), 2 mM L-glutamine (Gibco), 100 U/mL penicillin, and 100 μg/mL streptomycin (Capricorn Scientific). The cells were passaged using phosphate-buffered saline and trypsin (ThermoFisher Scientific). The SH-SY5Y, H1299, LN-18, and TE-671 cells were provided by the Heinrich Pette Institute for Experimental Virology (Hamburg, Germany); the SK-BR-3 cells were obtained from the collection of the Institute of Cytology RAS (St. Petersburg, Russia); the remaining cell cultures came from the collection of the Engelhardt Institute of Molecular Biology RAS (Moscow, Russia). All the cell lines were regularly tested for mycoplasma contamination every two weeks using Hoechst-33342 DNA staining (Sigma-Aldrich, USA).

All the inhibitors used in this study were dissolved in dimethyl sulfoxide (DMSO); stock solutions were stored at –80°C (*Table S1*).

RNA extraction

RNA for the transcriptome analysis was extracted from 1×10^6 SH-SY5Y cells treated with 5 μM 5-Aza for 24 h. RNA extraction was performed using the phenol–chloroform method with the TRIzol reagent (Ambion), followed by treatment with DNase

(Zymo Research, USA) and purification using RNA Clean & Concentrator-25 columns (Zymo Research), in accordance with the manufacturers' protocols.

The quantity of extracted RNA was measured using the Qubit 4 fluorometer (Thermo Fisher Scientific). Total RNA integrity was assessed using an Agilent 2100 bioanalyzer (Agilent Technologies, USA). The RNA integrity number (RIN) for each sample was $\geqslant 8$.

RNA sequencing and transcriptome analysis

A total of 1 μg of RNA was used to prepare each library. mRNA sequencing libraries were constructed using the TruSeq mRNA Library Prep Kit (Illumina, USA), in accordance with the manufacturer's instructions. Various single-index adapters from the TruSeq RNA Single Index kits (Illumina) were ligated to each sample to facilitate multiplex sequencing. DNA fragments 250–300 bp long were selected using MagPure A4 XP magnetic beads (Magen Biotechnology, China). The cDNA libraries were then enriched by PCR and purified. Library quality was assessed using the Agilent 2100 Bioanalyzer. Equimolar amounts of the final libraries were pooled and sequenced on the NextSeq 2000 platform (Illumina) in the single-end mode with a sequenced read length of 101 bp. The sequencing data were analyzed using the STAR aligner software, version 2.7.4a [20], in the "GeneCounts" mode, with the Ensembl human transcriptome annotation (GRCh38 assembly version; GRCh38.89 transcript annotation). Raw RNA-seq expression values (in the ReadsPerGene format) were normalized according to the DESeq2 standard [21]. Pathway activation levels (PALs) were calculated for a total of 3024 pathways using an open-access collection of molecular pathways retrieved from the Oncobox pathway databank [22].

Analysis of the activity of signaling pathways and CMAP analysis

The CMAP algorithm was employed to identify similar or opposite effects [23]. This algorithm allows to compare the changes in gene expression induced in response to specific treatments and each perturbation out of the hundreds of thousands cataloged in the database. The CMAP algorithm indicates which perturbation affected gene expression in a way most similar or opposite to the analyzed treatment. In this study, we compared the expression levels of the 100 genes most significantly upregulated and the 100 genes most significantly downregulated under the experimental conditions. DMSO-treated SH-SY5Y cells were used as controls.

Measurement of cell survival

Cell survival was measured using the Cell Proliferation Assay XTT kit (11465015001, Roche, Sigma-Aldrich, USA) and an AbiCell Resazurin Cytotoxicity Assay Kit (CEL-04-30ML, Abisense, Russia).

SH-SY5Y cells (2,500 cells per well in a 96-well plate) were co-incubated with the compounds for six days; the growth medium was then removed, and resazurin or the XTT reagent was added to the cells. After 4 hours of incubation at 37°C in the presence of 5% CO_2, the cell signaling level was measured using a Multiskan FC spectrophotometer by recording the difference in absorbance at 570 nm and 620 nm for resazurin, and at 450 nm and 605 nm for XTT. The changes in cell's survival ability caused by the cocktail of drugs was calculated as the difference between the total effect of the drugs and the sum of their individual effects. The method used to measure the cell survival and to calculate area under the curve (AUC) was similar to that described previously [24]. In order to calculate the AUC, the area under the cell survival (%) vs. drug concentration curve was determined by dividing the diagram into trapezoids. The AUC values calculated for all the cell lines were used to obtain the mean AUC value, which was then used for normalization. The AUC value was normalized so that a AUC equal to 1 corresponded to the mean AUC values across all the cell lines.

Cell counting on an automated microscope

Cell counting was performed on an automated fluorescence microscope using the protocol for detecting cells that express the ERK-KTR H2B-mRuby reporter system, which was previously utilized in our laboratory [25]. Each experiment was performed in three replicates: four random imaging fields were selected in each well to count cells. The cells were imaged at four time points: 0, 24, 72, and 144 h. The images of the cells were obtained using Leica DMI8 fluorescence microscope (Leica, Germany); cell counts were completed using the Cellpose and CellProfiler software.

Assessing the cell death mechanisms

SH-SY5Y, cells were seeded into 96-well plates at a density of 2500 cells per well. Staining was performed 72 h after the addition of 5-Aza. The following dyes were used to visualize mitochondria, tubulin, lysosomes, Fe^{II} ions, caspases 3/7, nuclei, and DNA: TMRE (Lumiprobe, Russia), Tubulin Tracker™ Deep Red (Invitrogen, USA), LumiTracker® LysoGreen (Lumiprobe), HMRhoNox-M (Lumiprobe), NucView® 488 (Biotium, USA), Hoechst-33342, and 7-aminoactinomycin D (7-AAD) (BioinnLabs, Russia), respectively. Staining was carried out at 37°C in an atmos-

phere of 5% CO_2. Imaging was performed using Leica DMI8 fluorescence microscope. *Supplementary Table S2* summarizes the concentrations, staining durations, and imaging parameters.

Four images of the cells co-incubated with the compound at each concentration were recorded; the experiment was performed in two replicates. In the images, individual cells were identified using the Cellpose and CellProfiler software. The protocols for cell segmentation and fluorescence intensity measurements had been published previously [24]. The activities of the mitochondria, lysosomes, and Fe^{II} ions were quantified using the integral fluorescence intensity of each cell. The percentage of stained cells was determined in the CellProfiler software to analyze the caspase 3/7 activity and identify dead cells by 7-AAD staining.

Confocal microscopy

The cells were fixed with a 4% formaldehyde solution (Sigma-Aldrich, USA) in 0.1 M phosphate-buffered saline (PBS) for 15 min and subsequently blocked using a solution containing 1% bovine serum albumin (BSA) (PanEco, Russia), 22.52 mg/mL glycine (Sigma-Aldrich), and 1% Tween (Sigma-Aldrich) in PBS. Alexa Fluor® 647-conjugated antibodies (ab194322, Abcam, UK) were utilized to study the distribution of TRK receptor proteins in the cytoplasm. The cells were co-incubated with antibodies overnight at a 1 : 100 antibody dilution in a 1% BSA solution in PBS. Coverslips were placed cell-side down onto glass slides containing 8 μL of the Slowfade gold medium (Invitrogen, USA) with 1 μg/mL DAPI (Sigma-Aldrich), and sealed with nail polish. Nuclei were visualized using DAPI staining. The data were obtained by confocal microscopy using Leica TCS SP5 laser scanning microscope (Leica) equipped with an HCX PLAPO CS 63×1.4 oil-immersion objective lens. The recorded confocal images (8-bit format) were analyzed in the LAS AF 4.0 software.

Data analysis

Statistical tests and data visualization were conducted using the GraphPad Prism 8.0, Python, and LAS AF software. The mean values and standard deviations (SD) or cell viability assessment were calculated using R and GraphPad Prism 8.0.

RESULTS

To assess the selectivity of 5-Aza toward neuroblastoma cells, we evaluated the effect of 5-Aza at concentrations ranging from 0.25 to 20 μM on various human cancer cell lines, including neuroblastoma SH-SY5Y, breast cancer SK-BR-3, renal cell carcinoma 786-O, cervical cancer SiHa, ovarian cancer SK-OV-3 colorectal carcinoma HCT-116, lung adenocarcinoma H1299, glioblastoma LN-18, and rhabdomyosarcoma TE-671 cells (*Fig. 1A*). The neuroblastoma SH-SY5Y cells were the ones most susceptible to 5-Aza (*Fig. 1B*).

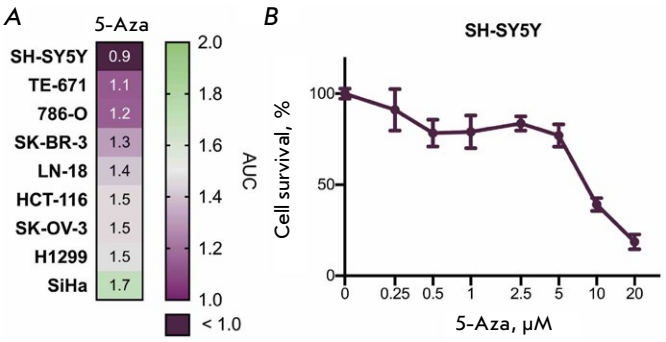


Fig. 1. Toxicity assessment of 5-azacitidine for human neuroblastoma SH-SY5Y cells. (*A*) Sensitivity of malignant cells of different origins to 5-azacitidine (5-Aza) within 72 h. The cells were treated with the drug at concentrations of 0.25–20 μM; the figure shows the AUC (area under the curve) values. (*B*) Survival of neuroblastoma SH-SY5Y cells after 5-Aza treatment for 72 h. The graphs show the average value of three replicates and the standard deviation (SD). Cells incubated with dimethyl sulfoxide (DMSO) were used as controls

Neuroblastoma SH-SY5Y cells were obtained by cloning a neuroblastoma SK-N-SH cell line [26]. SH-SY5Y is the cell line used in research most commonly: according to the data available at https://pubmed.ncbi.nlm.nih.gov/, the SH-SY5Y cell line was utilized in 13,789 publications, while the next most frequently used cell line, NMB, was mentioned in 5,338 publications. The SH-SY5Y cells harbor a mutation in the *ALK* gene (F1174L) [27] and are suitable for cell differentiation studies [28]. These cells exhibited the highest sensitivity to 5-Aza; so, further studies were performed using this cell culture.

In order to identify which cellular processes are affected by 5-Aza in neuroblastoma cells, we conducted a transcriptome analysis of the cells treated with 5 μM 5-Aza for 24 h and compared the findings to those for the transcriptome of SH-SY5Y cells exposed to DMSO for 24 h. The transcriptome analysis data are reported as signaling pathway activities and gene expression profiles (*Fig. 2A*).

The most prominent positive changes in signaling activity were observed in the pathways associated with transcription factors AP1, ATF6, and Sp1, as well as protein kinase B (PKB) activation; nerve growth factor (NGF) processing; cell cycle arrest mediated

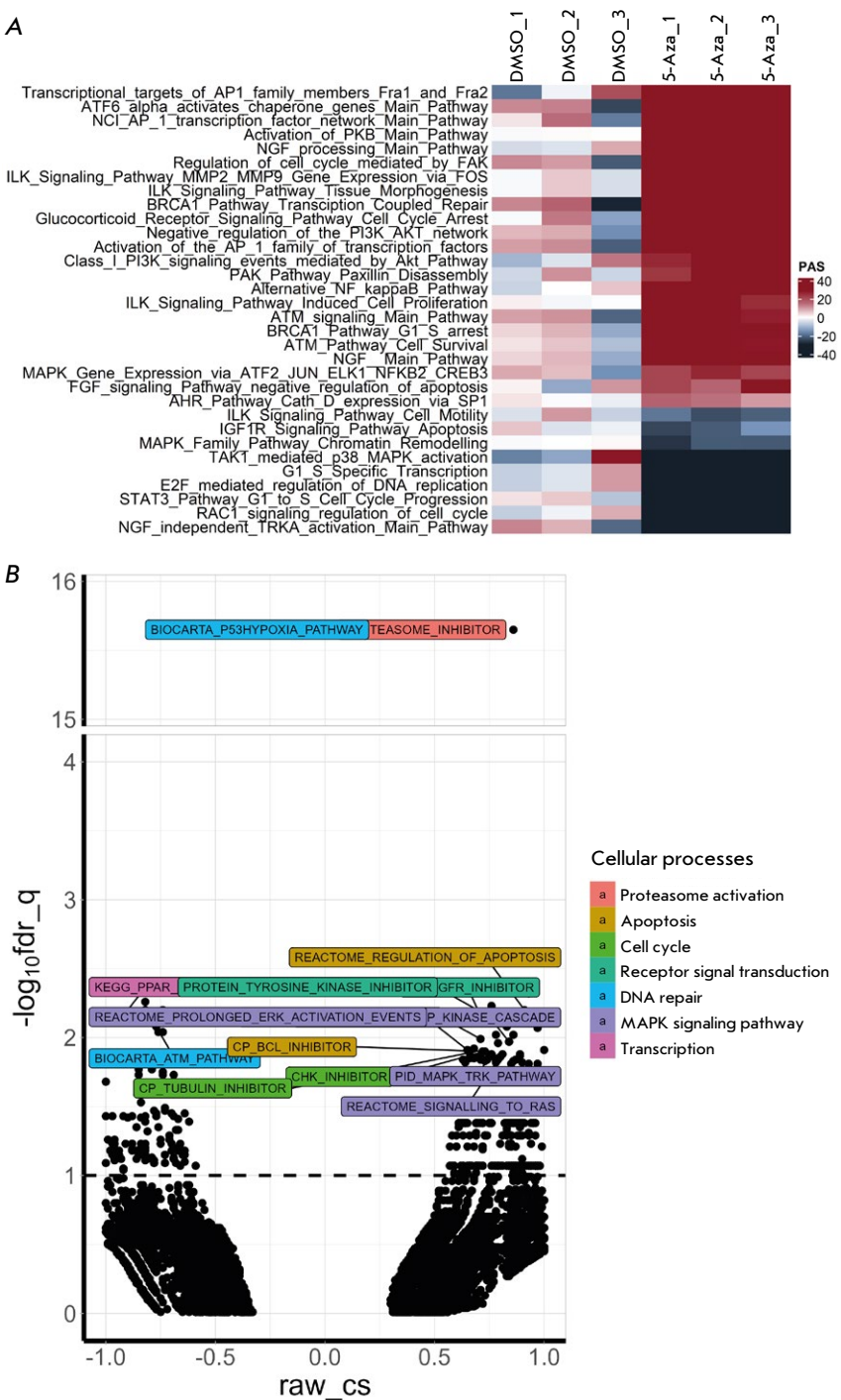


Fig. 2. Changes in the signaling pathways in human neuroblastoma SH-SY5Y cells after treatment with 5-azacitidine (5-Aza). (*A*) Pathway activation strength (PAS) in SH-SY5Y cells after treatment with 5 μM 5-Aza for 24 h according to the results of the Oncobox analysis [22]. The data are shown separately for each replicate. Pathway activation strength: the positive changes are shown in red; the negative changes are shown in black. The signaling pathways that may contribute to the progression of malignant tumors and incur statistically significant changes are shown. (*B*) Cellular processes that are altered in SH-SY5Y cells after treatment with 5 μM 5-Aza for 24 h according to the CMAP analysis. The dots indicate cellular processes from the CMAP analysis. Different colors indicate the classes of cellular processes with a reliable result according to the CMAP analysis. The results are shown as the inverse of the decimal logarithm of q-values after correction for multiple values ($-\log^{10}$fdr_q) and connectivity scores (raw_cs) and the degrees of similarity between differentially activated genes and the analyzed effect. Positive raw_cs values indicate identical changes in gene expression in response to treatment with 5-Aza and specified perturbations, while the negative values indicate opposite changes in gene expression in response to treatment with 5-Aza and specified perturbations

by glucocorticoid receptors; the integrin-linked kinase (ILK) mediated signaling pathway, and the PI3K/Akt/mTOR pathway (*Fig. 2A*). The most prominent negative changes were detected in the pathways associated with NGF-independent activation of receptor tropomyosin kinase A (TRKA), cell cycle, DNA replication involving the transcription factor E2F, the mitogen-activated protein kinase (MAPK) pathways, and apoptosis mediated by the insulin-like growth factor 1 receptor (IGF1R).

In order to determine what antitumor agents and cellular processes may exert a similar –or opposing – effect on the cellular transcriptomes, we conducted the CMAP analysis [23] (*Tables S3* and *S4; Figs. 2*

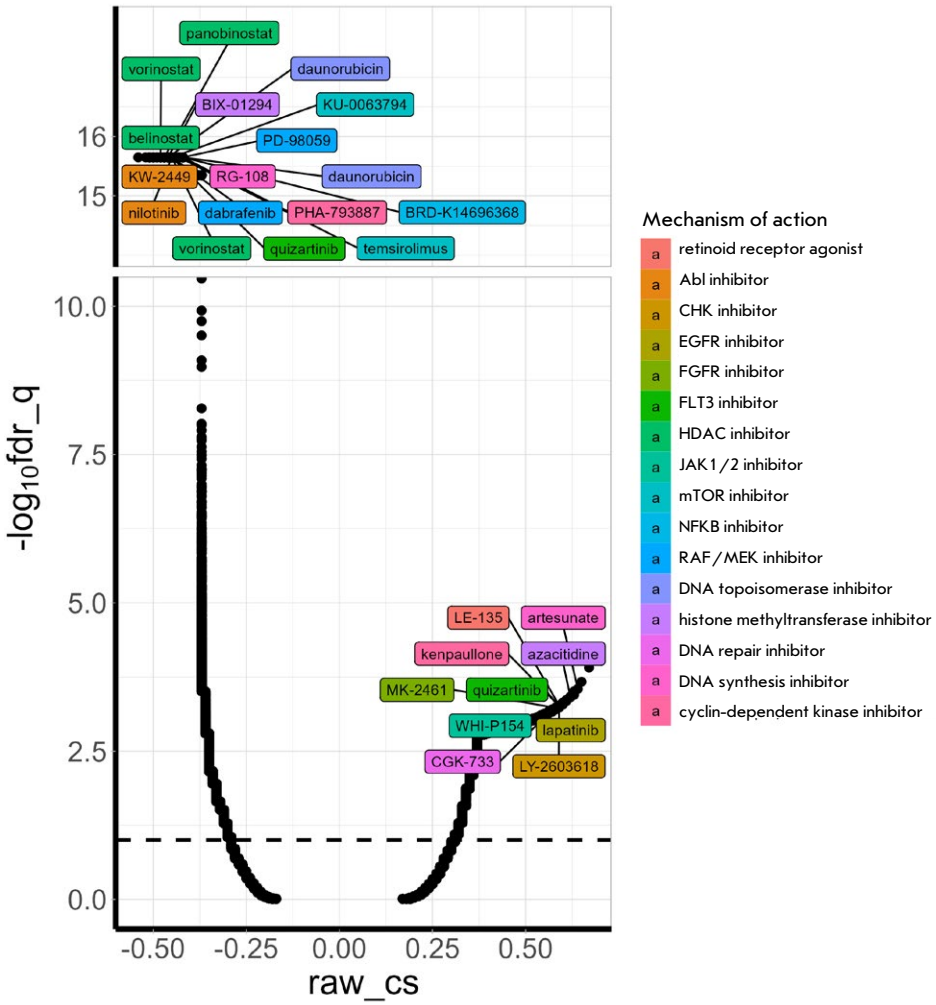


Fig. 3. Identifying drugs with an effect similar to that of 5-azacitidine (5-Aza) on the gene expression of human neuroblastoma SH-SY5Y cells using CMAP. The cells were treated with 5 μM 5-Aza and co-incubated with the drug for 24 h. The dots indicate the effects of inhibitors, small hairpin RNAs, or overexpression of certain genes. Different colors indicate the classes of inhibitors. Drugs with a statistically significant maximal effect are shown. The results are presented as the inverse of the decimal logarithm of q-values after correction for multiple values ($-\log^{10}$fdr_q) and connectivity scores (raw_cs) and the degrees of similarity between differentially activated genes and the analyzed effect. Positive raw_cs values indicate identical changes in gene expression in response to treatment with 5-Aza and specified perturbations, while the negative values indicate opposite changes in gene expression in response to treatment with 5-Aza and specified perturbations

and *3*). Our analysis revealed significant alterations in the cellular processes in SH-SY5Y cells treated with 5-Aza (*Fig. 2B*). Such alterations were primarily related to the processes associated with the regulation of apoptosis and the cell cycle, proteasomal activity, receptor signaling, and the MAPK pathway (in particular, those associated with extracellular signal-regulated kinase (ERK) and TRK). Opposing changes were observed for the processes related to the response to DNA damage and transcription.

Inhibitors of the epidermal growth factor receptor (EGFR), the fibroblast growth factor receptor (FGFR), Janus kinase (JAK), cell cycle checkpoint kinase (CHK), and DNA repair affected gene expression in a manner similar to that for 5-Aza (*Fig. 3*). 5-Aza was found to be one of the compounds eliciting comparable effects, attesting to the validity of the observed transcriptomic changes. Opposing effects were induced by inhibitors of histone deacetylases, mTOR, topoisomerase, RAF serine/threonine kinase, tyrosine kinase Abl, and the transcription factor NF-kB. Inhibitors of cyclin-dependent kinases (CDKs), receptor tyrosine kinase (RTK) FLT3, and DNA synthesis had different effects on gene expression.

We uncovered increased activities for seven apoptosis signaling pathways (*Table S3*). Since there exist several cell death mechanisms, we aimed to assess how 72 h treatment with 5-Aza would affect caspase 3/7, the mitochondria and lysosome activities, the Fe^{2+} content, and the number of dead SH-SY5Y cells. An up to 26% increase in the percentage of apoptotic cells was detected using fluorescent dyes (*Fig. 4A,B*), thus attesting to the enhanced activity of the apoptosis signaling pathways. The lysosomal activity in SH-SY5Y cells was increased after 5-Aza treatment (*Fig. 5A,B*). A slight decline in mitochondrial activity and increased Fe^{2+} levels were observed; however, these changes were minor and were likely to be related to cell death (*Fig. 5C,D*).

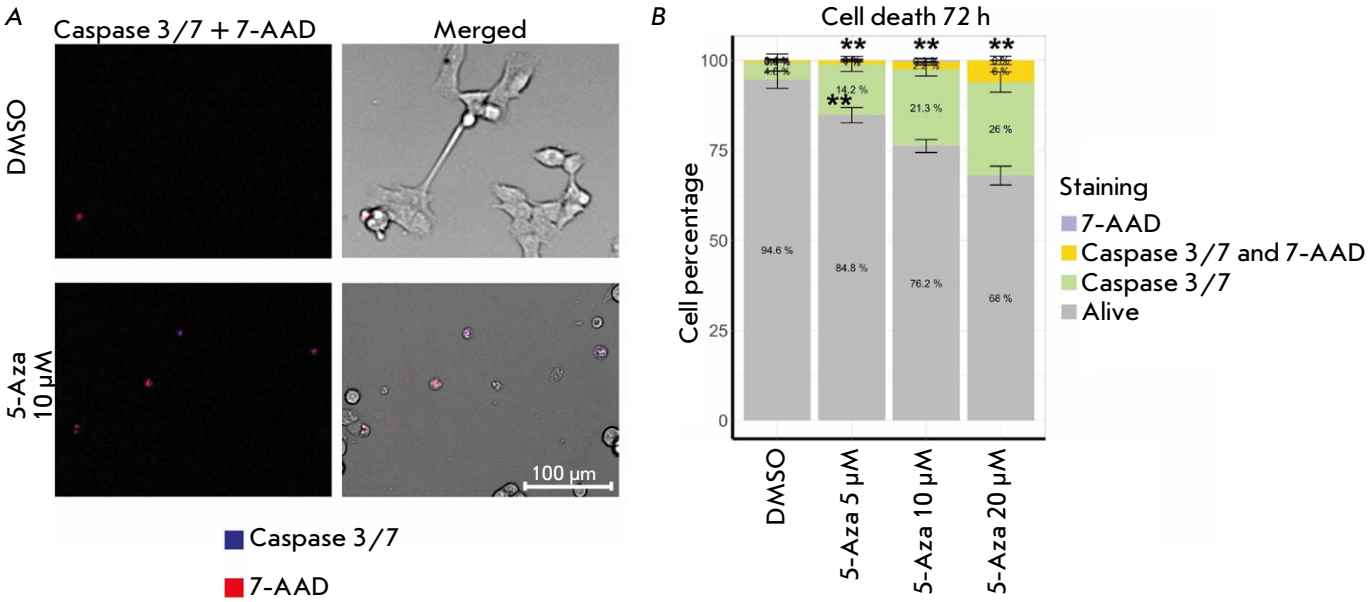


Fig. 4. The contribution of apoptosis to the death of human neuroblastoma SH-SY5Y cells after treatment with 5-azacitidine (5-Aza) for 72 h. (*A*) Caspase 3/7 and 7-aminoactinomycin D (7-AAD) staining in SH-SY5Y cells after treatment with 10 μM 5-Aza. (*B*) Apoptotic cells (green and yellow) in a population of SH-SY5Y cells after treatment with 5–20 μM 5-Aza. Cells were imaged using an automated fluorescence microscope. Cells co-incubated with dimethyl sulfoxide (DMSO) were used as control. The analysis was performed based on an assessment of the fluorescence intensity of dyes in 350–2,500 cells; the standard deviation (SD) was estimated for the average values for eight images for each 5-Aza concentration. Statistical significance was determined vs. DMSO using the Mann–Whitney U test (**$p \leq 0.01$)

In vitro studies have demonstrated that NGF can inhibit the proliferation of neurogenic cancer cell lines and induce their differentiation [29]. Since 5-Aza affects the activity of the signaling pathways mediated by NGF and its receptor TRKA, we assessed the distribution of TRK receptors within the cytoplasm of SH-SY5Y cells treated with 10 μM 5-Aza for 72 h (*Fig. 6*).

The observed increase in the intensity of the staining of SH-SY5Y cells with anti-TRK antibodies can explain the enhanced activity of the NGF- and TRK-mediated signaling pathways at the gene expression level.

Based on our findings (*Tables S3* and *S4*; *Figs. 2* and *3*), we selected 18 drugs that should be further tested, in combination with 5-Aza. In particular, we chose a number of inhibitors targeting RTK, histone deacetylases, the MAPK pathway, the cell cycle, as well as proteasomes, glucocorticoid receptors, DNA synthesis, DNA damage repair, apoptosis inducer, and the activator of the p38 signaling pathway (*Table 1*).

We simultaneously treated the cells with 2.5 μM 5-Aza and a second inhibitor at a pre-determined effective concentration that reduced cell viability by 20–50% within 72 h. The cells were subsequently incubated for another 144 h; cell viability was measured using a resazurin dye to assess the effectiveness of the drug combinations (*Fig. 7A*).

The two most effective combinations – 5-Aza with axitinib, a multikinase inhibitor, and mithramycin A, a DNA synthesis inhibitor – were identified. However, since this cell proliferation assessment method shows changes in the cellular metabolic activity [24], we tested the viability of the cells treated with combinations of inhibitors and 5-Aza by counting the cells on an automated fluorescence microscope (*Fig. 7B,C, S1*). We observed that the findings differed from those obtained in the experiment using resazurin dye: the highest effectiveness was attributable to the combinations of 5-Aza with mithramycin A and lonafarnib, a small G protein inhibitor (RAS) (*Fig. 7C*). Combinations of mithramycin A or lonafarnib with 5-Aza at concentrations that have no significant effect on cell proliferation led to substantial inhibition of SH-SY5Y cell growth and almost entirely stopped their proliferation.

DISCUSSION

We analyzed the changes in the transcriptome of human neuroblastoma SH-SY5Y cells treated with 5-Aza in a search for potential drugs that could be used in combination with 5-Aza. The mechanisms through

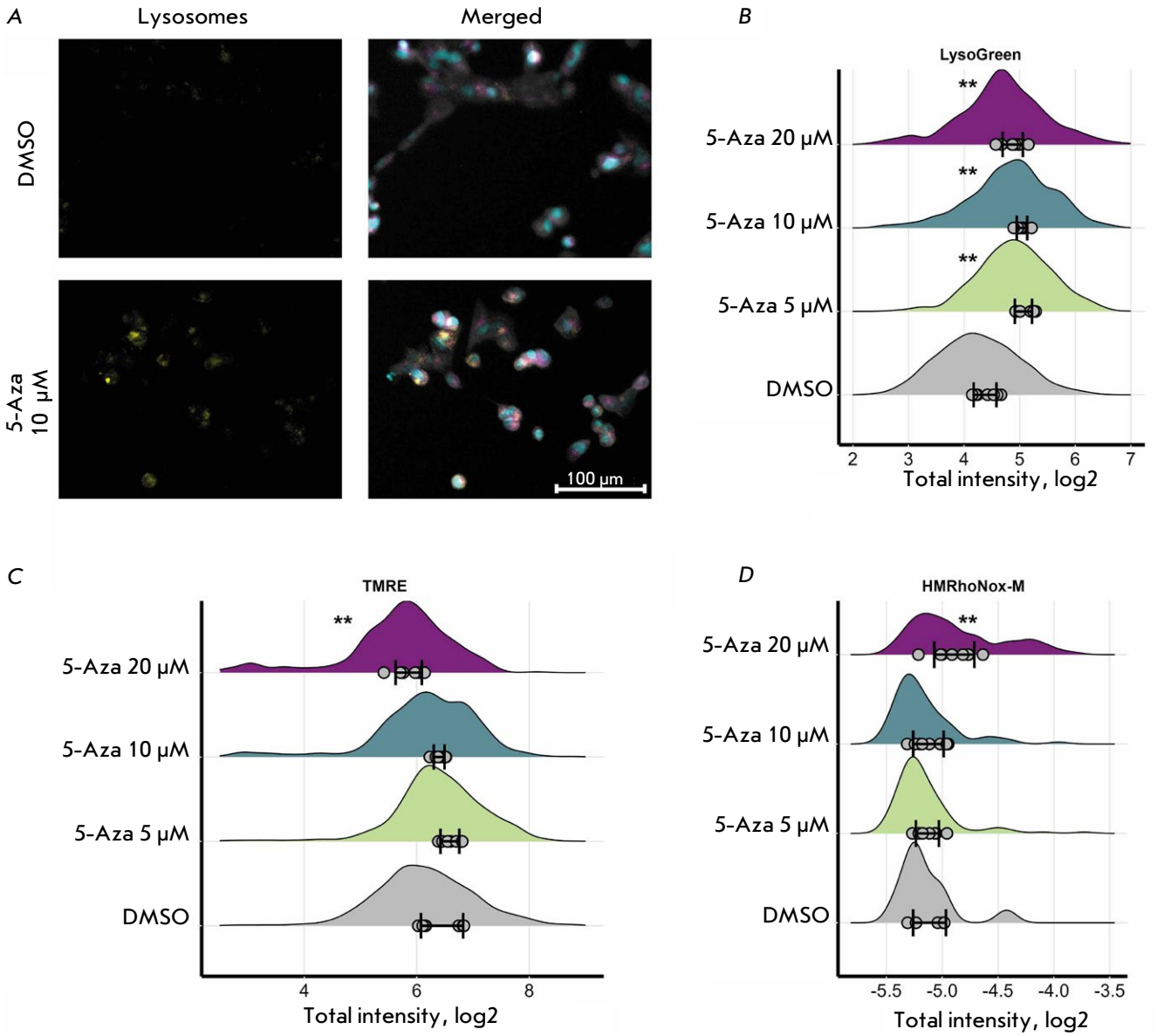


Fig. 5. Changes in the lysosomal activity in human neuroblastoma SH-SY5Y cells after treatment with 5-azacitidine (5-Aza) for 72 h. (*A*) Lysosome staining in SH-SY5Y cells after treatment with 10 μM 5-Aza. Cells were imaged using a fluorescence microscope. Lysosomes are shown in yellow; nuclei, in blue; mitochondria, in magenta; tubulin, in gray. (*B*) Changes in the lysosomal activity in SH-SY5Y cells after treatment with 5–20 μM 5-Aza. (*C*) Changes in the mitochondrial activity in SH-SY5Y cells after treatment with 5–20 μM 5-Aza. (*D*) Changes in the Fe^{II} iron content in SH-SY5Y cells after treatment with 5–20 μM 5-Aza. Cells co-incubated with dimethyl sulfoxide (DMSO) were used as controls. The distributions of the integrated dye intensity in 350–2500 cells are shown; the average values for each of the eight images are shown with dots; the standard deviation (SD) for the average values for the images is also indicated. Statistical significance was determined vs. DMSO using the Mann–Whitney U test (**$p \leq 0.01$)

which 5-Aza can induce cell death in SH-SY5Y cells were also identified.

5-Aza was found to affect the NGF-activated signaling pathways: it increases the intensity of cell staining using anti-TRK antibodies and alters the cell morphology. Earlier, it was demonstrated that 5-Aza induces the differentiation of neuroblastoma cells [30]. Among other factors, its action may have to do with the activation of cell differentiation. Some studies suggest that, in differentiated neuroblastoma cells, the absence of NGF induces apoptosis [31]. Differentiation induction by retinoic acid [32, 33] is extensively uti-

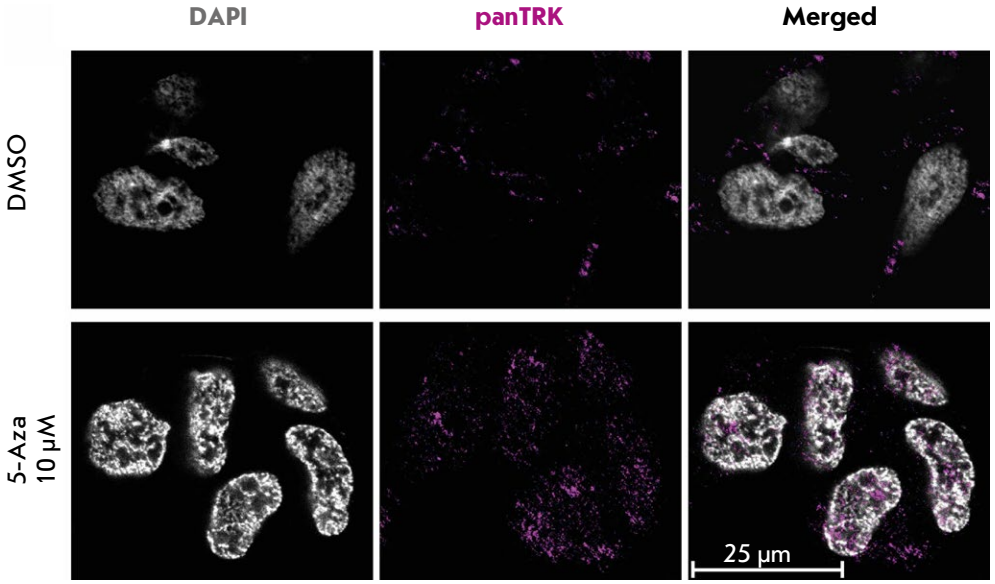


Fig. 6. The distribution of TRK proteins in the cytoplasm of human neuroblastoma SH-SY5Y cells after treatment with 10 μM 5-azacitidine (5-Aza) for 72 h. Cells co-incubated with dimethyl sulfoxide (DMSO) were used as controls. Cells were imaged by confocal microscopy using anti-TRK antibodies (Alexa647, magenta) and by staining the nuclei of fixed cells with DAPI (gray)

lized to treat low-risk neuroblastoma and as maintenance therapy for the more aggressive forms of the disease [34]. A combination of 5-Aza and retinoic acid was shown to enhance the differentiation of neuroblastoma cells [35]. 5-Aza increases caspase 3/7 activity, which may be associated with cell differentiation and upregulation of the TRK receptor expression in the absence of NGF.

We observed an increased lysosomal activity in SH-SY5Y cells. 5-Aza has also been shown to induce autophagy in acute myeloid leukemia cells [36]. A hypothesis has been put forward that 5-Aza can trigger various neuroblastoma cell death pathways; however, further research is needed to verify this hypothesis. A transcriptome analysis of SH-SY5Y cells revealed alterations in the pathways linked with mitochondria and cell death; nonetheless, we have detected no significant changes in mitochondrial activity.

Mithramycin A is an antibiotic active against lung, esophageal [37], colorectal cancer [38], as well as leukemia cells [39]; however, this drug has been found to be highly toxic [40]. Lonafarnib has been tested in combination with ALK inhibitors in ALK-mutant neuroblastoma cells both *in vitro* and *in vivo* [41]. It has been demonstrated that both mithramycin A and lonafarnib can reduce DNA methylation levels [42, 43]. The enhanced effectivity of 5-Aza in SH-SY5Y cells when used in combination with these drugs can potentially be mediated by their effects on DNA methyltransferase 1 (DNMT1). Furthermore, neuroblastoma cells are strongly dependent on the activation of certain growth factor receptors [44, 45]. 5-Aza triggers several RTK-mediated signaling pathways, while

Table 1. Selected drugs for assessing the effectiveness of their synergistic action on SH-SY5Y cells when used in combination with 5-azacitidine

Drug	Inhibitor class
Axitinib, entrectinib, gefitinib, sorafenib	Tyrosine kinase inhibitors
Belinostat, entinostat, vorinostat	Histone deacetylase inhibitors
Bortezomib	Proteasome inhibitor
Dexamethasone	Glucocorticoid receptor inhibitor, differentiation agent
Lonafarnib, PD184352	MAPK inhibitors
Mithramycin A	DNA synthesis inhibitor, inhibitor of Sp1 transcriptional activity
Metformin	Activator of the JNK/p38 MAPK pathway
BI2536, palbociclib, volasertib	Cell cycle inhibitors
Staurosporine	Apoptosis inducer, PKC inhibitor
Talazoparib	DNA damage response inhibitor

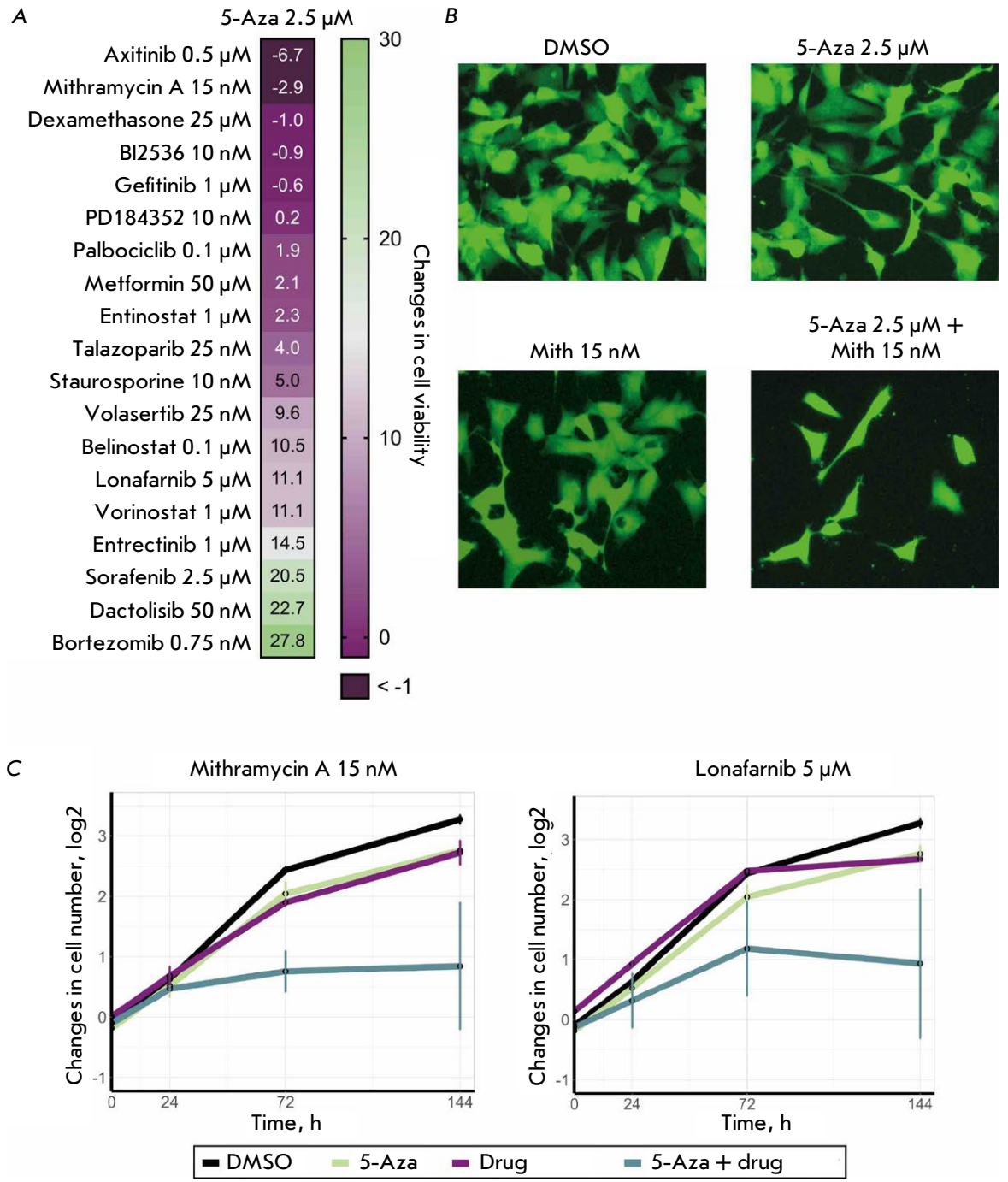


Fig. 7. The effectiveness of combinations of 5-azacitidine (5-Aza) and antitumor drugs against human neuroblastoma SH-SY5Y cells. The cells were simultaneously treated with 2.5 μM 5-Aza and an antitumor drug (the drugs and their concentrations are shown in the figure) and co-incubated for 144 h. Cells co-incubated with dimethyl sulfoxide (DMSO) were used as controls. (*A*) The heatmap showing the synergistic effect of a combination of 5-Aza and inhibitors belonging to different classes for SH-SY5Y cells. (*B*) Images of SH-SY5Y cells expressing the ERK-KTR H2B-Ruby reporter system after treatment with a combination of 2.5 μM 5-Aza and 15 nM mithramycin A (Mith) for 144 h. Cells were imaged by fluorescence microscopy. (*C*) The diagrams of changes in the number of SH-SY5Y cells after simultaneous addition of 2.5 μM 5-Aza and 15 nM mithramycin A or 5 μM lonafarnib. The diagrams show the average value of three replicates and the standard deviation (SD)

lonafarnib can block signal transduction from RTK by inhibiting RAS [41].

CONCLUSIONS

The analysis of the changes in the transcriptome of cells exposed to 5-Aza has identified drugs that exert a synergistic effect on neuroblastoma cell death and, in particular, the synergistic effect of a combination of 5-Aza and mithramycin A and lonafarnib against neuroblastoma SH-SY5Y cells. Further studies focusing on the effectiveness of drug combinations can pursue a more thorough analysis of the mechanism of the synergistic effect of these drugs and test the drug combinations in other neuroblastoma models. ●

This work was supported by the Russian Science Foundation (grant No. 22-14-00353). RNA sequencing was carried out using the equipment of the Core Facility "Genome" of the Institute of Molecular Biology, RAS (http://www.eimb.ru/rus/ckp/ccu_genome_c.php). The microscopy studies of genetically modified cells were supported by the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-15-2019-1660).

Supplementaries are available online at https://doi.org/10.32607/actanaturae.27558.

REFERENCES

1. Matthay K.K., Maris J.M., Schleiermacher G., Nakagawara A., Mackall C.L., Diller L., Weiss W.A. // Nat. Rev. Dis. Prim. 2016. V. 2. № 1. P. 16078. doi: 10.1038/nrdp.2016.78.
2. Irwin M.S., Naranjo A., Zhang F.F., Cohn S.L., London W.B., Gastier-Foster J.M., Ramirez N.C., Pfau R., Reshmi S., Wagner E., et al. // J. Clin. Oncol. 2021. V. 39. № 29. P. 3229–3241. doi: 10.1200/JCO.21.00278.
3. Krystal J., Foster J.H. // Children. 2023. V. 10. № 8. P. 1302. doi: 10.3390/children10081302.
4. Straathof K., Flutter B., Wallace R., Jain N., Loka T., Depani S., Wright G., Thomas S., Cheung G.W.-K., Gileadi T., et al. // Sci. Transl. Med. 2020. V. 12. № 571. P. eabd6169. doi: 10.1126/scitranslmed.abd6169.
5. Foster J.H., Voss S.D., Hall D.C., Minard C.G., Balis F.M., Wilner K., Berg S.L., Fox E., Adamson P.C., Blaney S.M., et al. // Clin. Cancer Res. 2021. V. 27. № 13. P. 3543–3548. doi: 10.1158/1078-0432.CCR-20-4224.
6. Fischer M., Moreno L., Ziegler D.S., Marshall L.V., Zwaan C.M., Irwin M.S., Casanova M., Sabado C., Wulff B., Stegert M., et al. // Lancet Oncol. 2021. V. 22. № 12. P. 1764–1776. doi: 10.1016/S1470-2045(21)00536-2.
7. Liu T., Merguerian M.D., Rowe S.P., Pratilas C.A., Chen A.R., Ladle B.H. // Cold Spring Harb. Mol. Case Stud. 2021. V. 7. № 4. P. 1–16. doi: 10.1101/MCS.A006064.
8. Chen L., Pastorino F., Berry P., Bonner J., Kirk C., Wood K.M., Thomas H.D., Zhao Y., Daga A., Veal G.J., et al. // Int. J. Cancer. 2019. V. 144. № 12. P. 3146–3159. doi: 10.1002/ijc.32058.
9. Geoerger B., Morland B., Jiménez I., Frappaz D., Pearson A.D.J., Vassal G., Maeda P., Kincaide J., Mueller U., Schlief S., et al. // Eur. J. Cancer. 2021. V. 153. P. 142–152. doi: 10.1016/j.ejca.2021.05.023.
10. Zage P.E. // Children. 2018. V. 5. № 11. P. 148. doi: 10.3390/children5110148.
11. Pieniążek B., Cencelewicz K., Bździuch P., Młynarczyk Ł., Lejman M., Zawitkowska J., Derwich K. // Int. J. Mol. Sci. 2024. V. 25. № 14. P. 7730. doi: 10.3390/ijms25147730.
12. Das P.M., Singal R. // J. Clin. Oncol. 2024. V. 22. № 22. P. 4632–4642. doi: 10.1200/JCO.2004.07.151.
13. Kaminskas E., Farrell A., Abraham S., Baird A., Hsieh L.-S., Lee S.-L., Leighton J.K., Patel H., Rahman A., Sridhara R., et al. // Clin. Cancer Res. 2005. V. 11. № 10. P. 3604–3608. doi: 10.1158/1078-0432.CCR-04-2135.
14. Steensma D.P. // Leuk. Res. 2009. V. 33. P. S12–S17. doi: 10.1016/S0145-2126(09)70228-0.
15. Hollenbach P.W., Nguyen A.N., Brady H., Williams M., Ning Y., Richard N., Krushel L., Aukerman S.L., Heise C., MacBeth K.J. // PLoS One. 2010. V. 5. № 2. P. e9001.
16. Gómez S., Castellano G., Mayol G., Suñol M., Queiros A., Bibikova M., Nazor K.L., Loring J.F., Lemos I., Rodríguez E., et al. // Epigenomics. 2015. V. 7. № 7. P. 1137–1153. doi: 10.2217/epi.15.49.
17. Jubierre L., Jiménez C., Rovira E., Soriano A., Sábado C., Gros L., Llort A., Hladun R., Roma J., de Toledo J.S., et al. // Exp. Mol. Med. 2018. V. 50. № 4. P. 1–12. doi: 10.1038/s12276-018-0077-2.
18. George R.E., Lahti J.M., Adamson P.C., Zhu K., Finkelstein D., Ingle A.M., Reid J.M., Krailo M., Neuberg D., Blaney S.M., et al. // Pediatr. Blood Cancer. 2010. V. 55. № 4. P. 629–638. doi: 10.1002/pbc.22607.
19. Vagapova E., Kozlov M., Lebedev T., Ivanenko K., Leonova O., Popenko V., Spirin P., Kochetkov S., Prassolov V. // Biomedicines. 2021. V. 9. № 12. P. 1846. doi: 10.3390/biomedicines9121846.
20. Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T.R. // Bioinformatics. 2013. V. 29. № 1. P. 15–21. doi: 10.1093/bioinformatics/bts635.
21. Love M.I., Huber W., Anders S. // Genome Biol. 2014. V. 15. P. 550. doi: 10.1186/s13059-014-0550-8.
22. Zolotovskaia M.A., Tkachev V.S., Guryanova A.A., Simonov A.M., Raevskiy M.M., Efimov V.V., Wang Y., Sekacheva M.I., Garazha A.V., Borisov N.M., et al. // Comput. Struct. Biotechnol. J. 2022. V. 20. P. 2280–2291. doi: 10.1016/j.csbj.2022.05.006.
23. Subramanian A., Narayan R., Corsello S.M., Peck D.D., Natoli T.E., Lu X., Gould J., Davis J.F., Tubelli A.A., Asiedu J.K., et al. // Cell. 2017. V. 171. № 6. P. 1437–1452. e17. doi: 10.1016/j.cell.2017.10.049.
24. Mikheeva A., Bogomolov M., Gasca V., Sementsov M., Spirin P., Prassolov V., Lebedev T. // Cell Death Discov. 2024. V. 10. P. 181. doi: 10.1038/s41420-024-01950-3.
25. Lebedev T.D., Khabusheva E.R., Mareeva S.R., Ivanenko K.A., Morozov A.V., Spirin P.V., Rubtsov P.M., Snezhkina A.V., Kudryavtseva A.V., Sorokin M.I., et al. // J. Biol. Chem. 2022. V. 298. № 8. P. 102226. doi: 10.1016/j.

jbc.2022.102226.
26. Biedler J.L., Roffler-Tarlov S., Schachner M., Freedman L.S. // Cancer Res. 1978. V. 38. № 11. Part 1. P. 3751–3757.
27. George R.E., Sanda T., Hanna M., Fröhling S., Luther W. 2nd, Zhang J., Ahn Y., Zhou W., London W.B., McGrady P., et al. // Nature. 2008. V. 455. № 7215. P. 975–978. doi: 10.1038/nature07397.
28. Kovalevich J., Langford D. // Methods Mol. Biol. 2013. V. 1078. P. 9–21. doi: 10.1007/978-1-62703-640-5_2.
29. Aloe L., Rocco M.L., Balzamino B.O., Micera A. // J. Exp. Clin. Cancer Res. 2016. V. 35. № 1. P. 1–7. doi: 10.1186/s13046-016-0395-y.
30. Bartolucci S., Rossi M., Longo A., Rossi M., Estenoz M., Momparler R.L., Santoro B., Augusti-Tocco G. // Cell Differ. Dev. 1989. V. 27. № 1. P. 47–55. doi: 10.1016/0922-3371(89)90043-9.
31. Nakagawara A., Arima-Nakagawara M., Scavarda N.J., Azar C.G., Cantor A.B., Brodeur G.M. // N. Engl. J. Med. 1993. V. 328. № 12. P. 847–854. doi: 10.1056/NEJM199303253281205.
32. Bayeva N., Coll E., Piskareva O. // J. Pers. Med. 2021. V. 11. № 3. P. 211. doi: 10.3390/jpm11030211.
33. Lebedev T.D., Vagapova E.R., Prassolov V.S. // Acta Naturae. 2021. V. 13. № 4. P. 69–77. doi: 10.32607/actanaturae.11461.
34. Makimoto A., Fujisaki H., Matsumoto K., Takahashi Y., Cho Y., Morikawa Y., Yuza Y., Tajiri T., Iehara T. // Cancers (Basel). 2024. V. 16. № 3. P. 544. doi: 10.3390/cancers16030544.
35. Almeida V.R., Vieira I.A., Buendia M., Brunetto A.T., Gregianin L.J., Brunetto A.L., Klamt F., de Farias C.B., Abujamra A.L., Lopez P.L. da C., et al. // Mol. Neurobiol. 2017. V. 54. № 10. P. 7610–7619. doi: 10.1007/s12035-016-0250-3.
36. Noronha N., Durette C., Cahuzac M., E Silva B., Courtois J., Humeau J., Sauvat A., Hardy M.-P., Vincent K., Laverdure J.-P., et al. // Leukemia. 2024. V. 38. № 5. P. 1019–1031. doi: 10.1038/s41375-024-02250-6.
37. Zhang M., Mathur A., Zhang Y., Xi S., Atay S., Hong J.A., Datrice N., Upham T., Kemp C.D., Ripley R.T., et al. // Cancer Res. 2012. V. 72. № 16. P. 4178–4192. doi: 10.1158/0008-5472.CAN-11-3983.
38. Quarni W., Dutta R., Green R., Katiri S., Patel B., Mohapatra S.S., Mohapatra S. // Sci. Rep. 2019. V. 9. № 1. P. 15202. doi: 10.1038/s41598-019-50917-3.
39. Vagapova E.R., Lebedev T.D., Tikhonova A.D, Goikhman B.V., Ivanenko K.A., Spirin P.V., Prassolov V.S. // Mol. Biol. 2020. V. 54. № 3. P. 458–463. doi: 10.1134/S002689332003019X.
40. Baum M. // Br. J. Cancer. 1968. V. 22. № 2. P. 176–183. doi: 10.1038/bjc.1968.25.
41. Pucci P., Lee L.C., Han M., Matthews J.D., Jahangiri L., Schlederer M., Manners E., Sorby-Adams A., Kaggie J., Trigg R.M., et al. // Nat. Commun. 2024. V. 15. P. 3422. doi: 10.1038/s41467-024-47771-x.
42. Lin R., Hsu C.-H., Wang Y.-C. // Anticancer. Drugs. 2007. V. 18. № 10. P. 1157–1164. doi: 10.1097/CAD.0b013e3282a215e9.
43. Chen T., Cai C., Wang L., Li S., Chen L. // Front. Pharmacol. 2020. V. 11. P. 589780.
44. Lebedev T., Buzdin A., Khabusheva E., Spirin P., Suntsova M., Sorokin M., Popenko V., Rubtsov P., Prassolov V. // Int. J. Mol. Sci. 2022. V. 23. № 14. P. 7724. doi: 10.3390/ijms23147724.
45. Lebedev T., Vagapova E., Spirin P., Rubtsov P., Astashkova O., Mikheeva A., Sorokin M., Vladimirova U., Suntsova M., Konovalov D., et al. // Oncogene. 2021. V. 40. № 44. P. 6258–6272. doi: 10.1038/s41388-021-02018-7.

Integration of HiMoRNA and RNA-Chrom: Validation of the Functional Role of Long Non-coding RNAs in the Epigenetic Regulation of Human Genes Using RNA-Chromatin Interactome Data

I. S. Ilnitskiy[1,2,3†*], G. K. Ryabykh[1,2†], D. A. Marakulina[3,4], A. A. Mironov[1,2], Yu. A. Medvedeva[3,4]
[1]Lomonosov Moscow State University, Faculty of Bioengineering and Bioinformatics, Moscow, 119234 Russia
[2]Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, 119991 Russia
[3]Skryabin Institute of Bioengineering, FRC Biotechnology, Russian Academy of Sciences, Moscow, 117312 Russia
[4]School of Biomedical Physics, Moscow Institute of Physics and Technology, Dolgoprudny, Moscow, 141701 Russia
*E-mail: nfsus96@gmail.com
†Authors contributed equally to the work.
Received: October 21, 2024; in final form, March 31, 2025
DOI: 10.32607/actanaturae.27543

ABSTRACT **Long non-coding RNAs (lncRNAs) play a crucial role in the epigenetic regulation of gene expression by recruiting chromatin-modifying proteins to specific genomic loci. Two databases, previously developed by our groups, HiMoRNA and RNA-Chrom, provide valuable insights into this process. The former contains data on epigenetic modification regions (peaks) correlated with lncRNA expression, while the latter offers genome-wide RNA–chromatin interaction data for tens of thousands of RNAs. This study integrated the two resources to generate experimentally supported, interpretable hypotheses regarding lncRNA-mediated epigenetic gene expression regulation. We adapted the web interfaces of HiMoRNA and RNA-Chrom to enable the retrieval of chromatin contacts for each "lncRNA–epigenetic modification–associated gene" triad from HiMoRNA, either at specific genomic loci or genome-wide via RNA-Chrom. The integration analysis revealed that for the lncRNAs MALAT1, HOXC-AS2, NEAT1, NR2F1-AS1, PVT1, and MEG3, most HiMoRNA peaks are located within 25 kb of their RNA-Chrom contacts. Further investigation confirmed the RNA–chromatin contacts of MIR31HG and PVT1 lncRNAs, with HiMoRNA peaks for H3K27ac and H3K27me3 marks in the loci of the genes *GLI2* and *LATS2*, respectively, which are known to be regulated by these RNAs. Thus, the integration of HiMoRNA and RNA-Chrom offers a powerful platform to elucidate the role of specific lncRNAs in the regulation of histone modifications at both individual loci and genome-wide levels. We expect this integration to help significantly advance the functional annotation of human lncRNAs.**
KEYWORDS **long non-coding RNA, histone modification, RNA-chromatin interaction.**
ABBREVIATIONS **ncRNA – non-coding RNA; lncRNA – long non-coding RNA.**

INTRODUCTION

Human cells transcribe a vast number of long non-coding RNAs (lncRNAs), with their quantity comparable to that of protein-coding genes [1, 2]. Functional annotation of lncRNAs is challenging due to their low expression levels, tissue specificity, and low sequence conservation [3–5]. Nevertheless, lncRNAs were observed to preserve certain characteristics, notably synteny with neighboring genes, secondary structure, and similarity in short sequence fragments [6]. In addition, the transcriptional regulation of lncRNA transcription exhibits intricacy comparable to that of protein-coding RNAs, facilitating their involvement in diverse molecular processes [7]. Suppression of lncRNAs has been shown to result in significant changes in the transcriptional profile of

cells [8]. These findings indicate a functional role for numerous lncRNAs. Most lncRNAs were shown to interact with chromatin and to be involved in the epigenetic regulation of genomic loci and the structural organization of chromosomes [9–12]. Accordingly, identifying the functional genomic targets of chromatin-associated lncRNAs is of high value.

Previously, we developed the HiMoRNA database [13], which catalogues over 5 million epigenetic "peaks," genomic regions exhibiting one of 10 histone modifications, with their modification level significantly correlating with lncRNA expression across at least 20 cell lines and tissues. Where possible, histone modification peaks in HiMoRNA are associated with genes, forming a triad: "lncRNA–epigenetic modification peak–associated gene." Within such a triad, the lncRNA is hypothesized to modulate the expression of the corresponding gene through the promotion or repression of histone modification within the gene's associated peak region. In the case of promotion, peaks positively correlate with lncRNA expression ("+" peaks), while in the case of repression, the correlation is negative ("–" peaks). These associations enable the formulation of hypotheses regarding the role of lncRNAs in the modulation of epigenetic modifications at specific genomic loci and the regulation of gene expression. However, to formulate and empirically test reasonable hypotheses, the 5-million-peak dataset has to be pre-processed to select the most reliable peaks. Experimental methods for detecting RNA-chromatin interactions can provide valuable data for this purpose. Several experimental methods exist to identify chromatin regions interacting with non-coding RNAs. These methods can be broadly classified into two categories: "one-to-all" [11, 14–18], which identify the contacts of a specific RNA with chromatin, and "all-to-all" [19–24], which capture all possible RNA–DNA contacts in a cell [25]. Notwithstanding their utility, these approaches are prone to high false positive rates. Additionally, "all-to-all" methods show low sensitivity to lowly expressed RNAs and bias toward nascent transcripts. Despite these challenges, genome-wide data on non-coding RNA (ncRNA) interactions with chromatin are crucial for elucidating their mechanisms of action. In this context, the RNA-Chrom database [26] was recently created. It contains experimental data on thousands of RNA-chromatin contacts and offers two analytical modes ("from RNA" and "from DNA") that can be used for research purposes.

To improve and streamline the functional annotation of lncRNAs, we have integrated the HiMoRNA and RNA-Chrom databases. The web interfaces of HiMoRNA and RNA-Chrom were modified to provide direct access to chromatin contacts for 4,124 out of the 4,145 lncRNAs from HiMoRNA in RNA-Chrom. This integration enables the generation of hypotheses regarding the mechanisms of epigenetic regulation of human gene expression by long non-coding RNAs, supported by experimental data on their interactions with chromatin. We anticipate that this unified resource will prove a valuable tool for identifying high-confidence "lncRNA–epigenetic modification–associated gene" triads for further experimental investigation of lncRNA mechanisms in gene regulation. The HiMoRNA database is available to users at https://himorna.fbras.ru (as of 20.10.2024).

EXPERIMENTAL PART

Integration of the HiMoRNA and RNA-Chrom databases

Due to differences in gene annotation sources (HiMoRNA uses GENCODE v31, while RNA-Chrom uses GENCODE v35), we established gene correspondence using three similarity metrics: (1) matching gene names ('gene_name', *Fig. 1A*); (2) matching gene identifiers ('gene_id', *Fig. 1A*); and (3) a Jaccard index (the ratio of the length of gene overlap to the length of their union) greater than 0.99 (Jaccard index > 0.99, *Fig. 1A*). Due to differences in naming conventions and genomic coordinates between annotation versions, gene identifiers and positions often do not align directly. To address this issue, we intersected 4,145 lncRNA genes from HiMoRNA with 60,619 genes from RNA-Chrom based on genomic coordinates using the "*intersect*" command from "*bedtools*". This yielded 6,778 gene pairs, exceeding the number of HiMoRNA entries, because some HiMoRNA genes intersected multiple times with RNA-Chrom genes. The two HiMoRNA genes (ENSG00000267034.1 and ENSG00000280076.1) did not intersect with any RNA-Chrom genes. Subsequently, the Jaccard index was calculated for each gene pair. The gene pairs were classified into six groups based on similarity metrics (*Fig. 1A*). By using a Jaccard index > 0.99 as the primary similarity metric, we identified four groups (groups 2, 4, 5, and 6) as having no overlapping genes (*Fig. 1B*), establishing 4,100 unambiguous gene correspondences. For the remaining 43 genes from HiMoRNA, 24 additional matches to RNA-Chrom genes were established using the 'gene_name' matching metric. In total, we identified 4,124 lncRNA genes common to both databases (see *Supplementary Table 1* for a full correspondence table that is available for download on the HiMoRNA web resource).

To support database integration and streamline access to chromatin interaction data, we made sever-

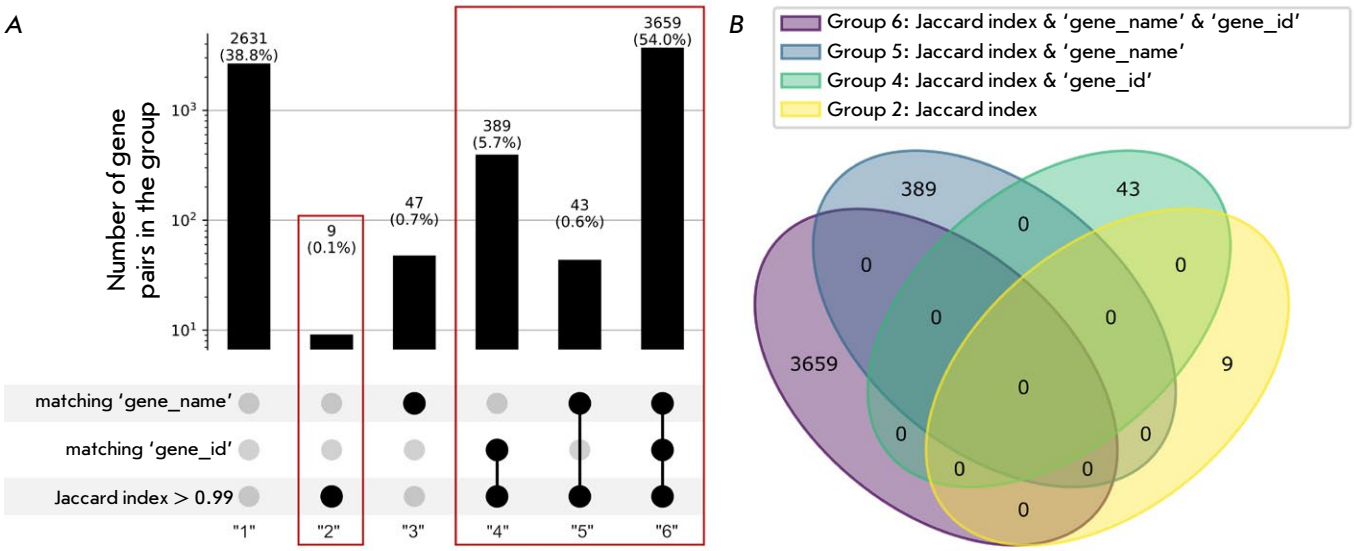


Fig. 1. Intersection of 4,145 genes from HiMoRNA with 60,619 genes from RNA-Chrom. (*A*) Division of gene pairs into six groups based on the similarity metrics they satisfy. Groups achieving unambiguous gene correspondence are highlighted with red rectangles. (*B*) Venn diagram showing the overlap between the gene groups 2, 4, 5, and 6 (the total number of gene pairs in these four groups is 4,100)

al enhancements to the RNA-Chrom and HiMoRNA interfaces. These include parameter processing (locus, RNA name, RNA-Chrom internal RNA identifier, organism) from a specific type of URL (e.g., https://rnachrom2.bioinf.fbb.msu.ru/basic_graphical_summary_dna_filter?locus=chrX:23456-24253566&name=XIST&rnaID=227896&organism=Homo+sapiens) and providing information on the chromatin contacts of the requested lncRNA across different experiment types on a new browser page.

Enhancements to the HiMoRNA interface include the incorporation of the gene correspondence table between RNA-Chrom and HiMoRNA to ensure correct URL generation.

A "Go to RNA-Chrom DB" button with a dropdown menu (*Fig. 2*) was added to the "Search Results Page," allowing the user to generate three types of URL links to navigate to the RNA-Chrom page:

a) Contacts of the given lncRNA in a specific genomic locus, extended by 1 / 5 / 10 / 25 / 50 / 100 kb;

b) All contacts of the given lncRNA;

c) All RNAs with contacts in a specific genomic locus.

One-sided Fisher's exact test

In most triads, histone modification peaks show both negative and positive correlations between lncRNA expression and the peak signal level (hereinafter denoted "–" and "+" peaks, respectively). The presence of a "+" peak suggests a role for the lncRNA in establishing the histone modification, whereas a "–" peak implies its role in removing the modification.

The alignment of predictions with published experimental results was evaluated by selecting lncRNAs and their corresponding histone peaks (positively and negatively correlated), extended by ±25 kb, and filtering for those where the proportion of peaks supported by contacts for at least one histone mark exceeded 40%. We subsequently conducted independent right-tailed and left-tailed Fisher's exact tests for each lncRNA-histone mark pair. For example, a representative contingency table for the "PVT1–H3K27ac" pair is provided in *Supplementary Table 2*.

Red-ChIP data

As a case study, we used the lncRNA PVT1 to validate integration between HiMoRNA and RNA-Chrom using independent experimental data. Specifically, we incorporated Red-ChIP data [27], available in the Gene Expression Omnibus under accession number GSE174474, samples GSM5315228 and GSM5315229 (hES cell line). The Red-ChIP method captures RNA-chromatin contacts mediated by the EZH2 protein, a component of the PRC2 complex, which establishes, among others, the H3K27me3 histone modification.

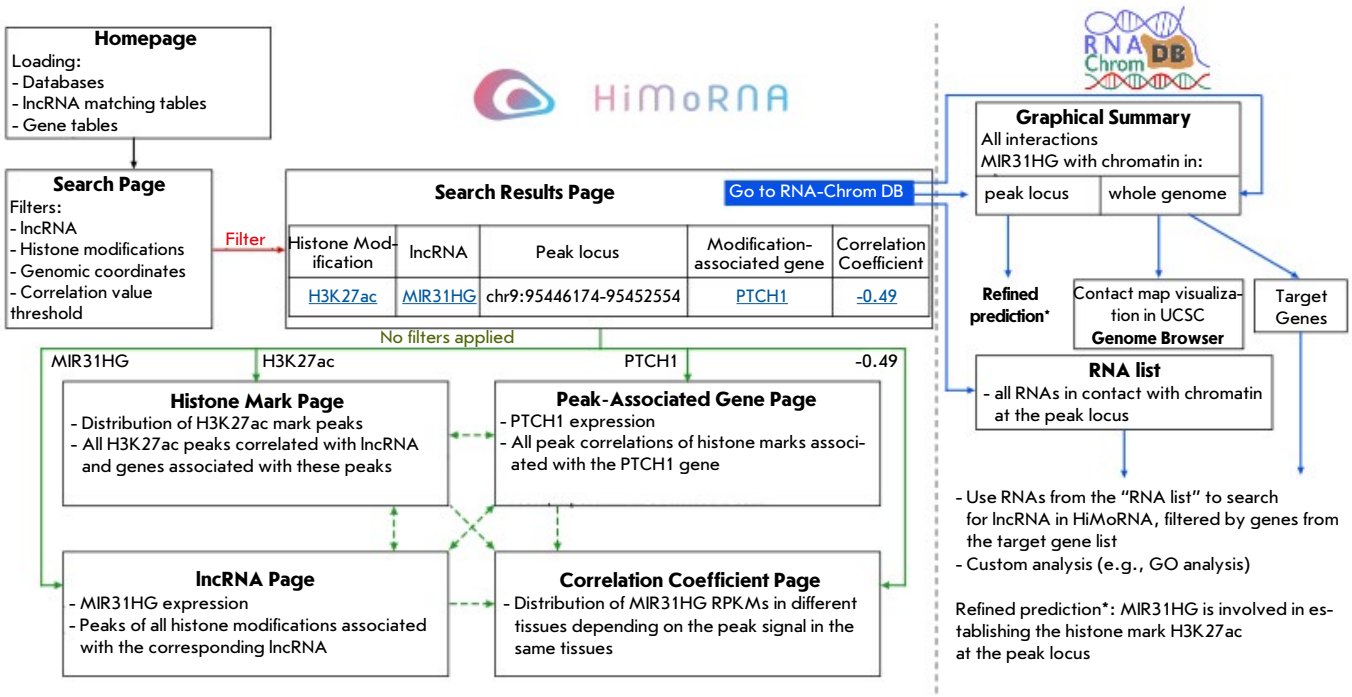


Fig. 2. Usage scenario of the HiMoRNA and RNA-Chrom databases after integration. Rectangles represent web pages, and arrows indicate movement between them

Primary data processing followed the established RNA-Chrom database protocol. Primary data processing followed the established RNA-Chrom database protocol. Subsequently, we identified genomic regions enriched with lncRNA PVT1 chromatin contacts using the BaRDIC program (--qval_type all; --qval_threshold 1) [28]. Consequently, 3,242 genomic regions exhibiting potential functionality and EZH2-mediated PVT1 binding were identified.

RESULTS

Integration of Databases

Given that HiMoRNA contains millions of epigenetic peaks, selecting the most reliable ones for further analysis is critical. For this task, we integrated HiMoRNA peak data with RNA-chromatin interactome data from the RNA-Chrom database. This integration was achieved by establishing one-to-one correspondences between genes in both databases and modifying their web interfaces (see "Experimental Section," subsection "Integration of HiMoRNA and RNA-Chrom Databases"). This strategy allows HiMoRNA to generate specific URL queries for 4,124 out of the 4,145 lncRNAs listed in RNA-Chrom, enabling, in particular, the identification of other chromatin loci with which the investigated RNA interacts. This approach significantly expands our understanding of the function of a specific RNA.

The general integration scheme is presented in *Fig. 2*. To utilize the integration, the user first needs to find the target lncRNA in the HiMoRNA database. From the HiMoRNA homepage, users can download the database itself, as well as the "Gene Table" and the "lncRNA Correspondence Table" added as part of the integration, to search for genes/lncRNAs of interest by genomic coordinates. This feature accommodates potential mismatches between user-specified Ensembl identifiers/lncRNA names and those present in the HiMoRNA database. On the search page, the user must configure filters according to their task, specifying the lncRNAs, histone modifications, genomic coordinates, and genes associated with the selected histone modifications.

On the search results page, users may conduct a more thorough examination of the retrieved predictions, including accessing the RNA-Chrom database. To do this, the user should select the desired

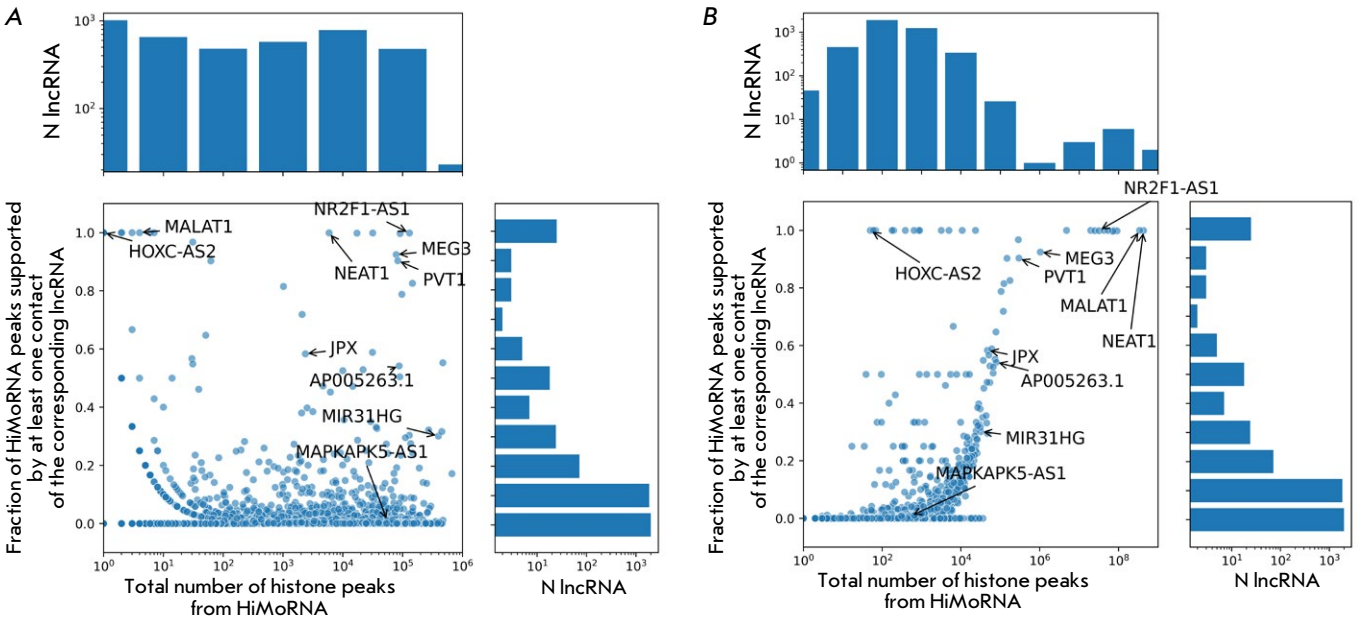


Fig. 3. Proportions of HiMoRNA peaks confirmed by at least one contact of the corresponding lncRNA from RNA-Chrom, relative to the total number of HiMoRNA peaks for the corresponding lncRNA (*A*) and the total number of contacts for the corresponding lncRNA from RNA-Chrom (*B*). Genomic coordinates of contacts are extended by ± 25 kb

"lncRNA–epigenetic modification peak–associated gene" triad from the interactive results table and then click the "Go to RNA-Chrom DB" button. In the dropdown menu, the user should select the appropriate option to navigate to a page displaying: 1) contacts of the given lncRNA in the region of a specific peak (with the option to choose how much to extend the peak coordinates when searching for contacts); 2) all contacts of the given lncRNA; or 3) all lncRNAs with contacts in the specified locus. The user will be automatically redirected to a graphic summary of the lncRNA-chromatin interactome on the RNA-Chrom database webpage. This summary allows the user to determine whether the functional relationship "lncRNA–epigenetic modification" from HiMoRNA is mediated by the physical presence of the lncRNA at the corresponding genomic locus, as well as to identify other lncRNAs potentially involved in the regulation of that locus. Visual analysis is enabled by utilizing the UCSC Genome Browser to load data from all relevant experiments by clicking "VIEW IN GENOME BROWSER." By selecting a single RNA-chromatin interactome experiment, the user can obtain a list of the genes located in the genomic region of interest, along with statistics on their interactions with the lncRNA, by clicking "ALL TARGET GENES." This list of genes can be downloaded for further research, such as performing a GO analysis. The "Use Cases" section offers a comprehensive discussion and exemplification of the HiMoRNA and RNA-Chrom database integration.

Consistency of HiMoRNA and RNA-Chrom Results

To assess the completeness of the integration, we analyzed the frequency of confirmation of histone peaks from HiMoRNA, correlated with lncRNA expression, using data on the corresponding lncRNA chromatin contacts from RNA-Chrom. Out of the 4,145 lncRNAs present in HiMoRNA, 4,011 (96.8%) were found to have at least one contact in the RNA-Chrom database, with 29 RNAs not matching between the databases and 105 (2.5%) having no contacts in RNA-Chrom. Among the 4,011 lncRNAs of interest, only 35.5% had at least one peak supported by the contacts of the corresponding lncRNA. However, due to the design of experimental protocols, actual lncRNA-chromatin interactions may occur at a distance from the experimentally detected contact. To address this issue, we extended the contact coordinates for a more accurate assessment of the correspondence between HiMoRNA predicted peaks and RNA-Chrom data. Extending contact ranges by ±1, ±5, ±10, ±25, and ±50 kb resulted in a respective increase in the per-

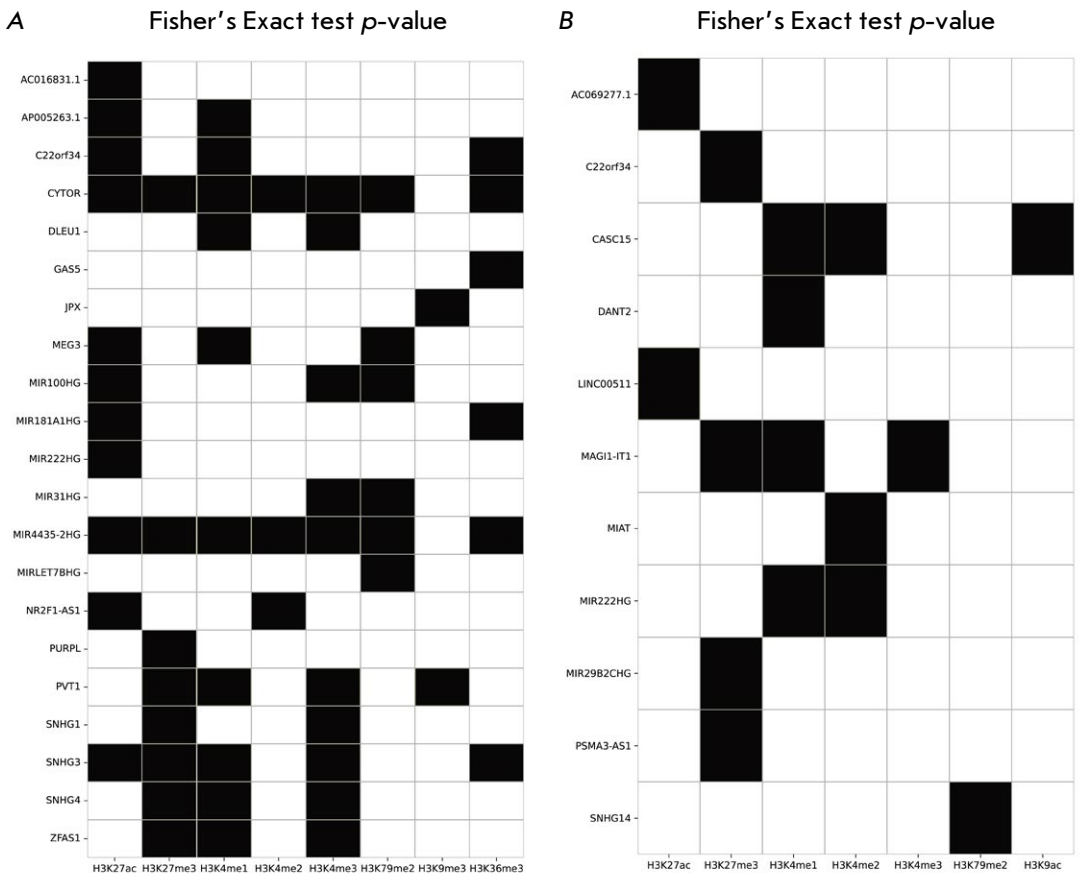


Fig. 4. Heatmap showing the results of Fisher's exact test for pairs of "lncRNA – histone mark peaks extended by ±25 kb". Black indicates that the proportion of "–" or "+" histone peaks supported by contacts of the corresponding lncRNA is greater than 0.4 and the p-value of the Fisher's exact test is less than 10^{-3}; otherwise, it is white. (*A*) Right-tailed Fisher's test: "+" peaks of corresponding histone marks are better supported by RNA-Chrom contacts than "–" peaks. (*B*) Left-tailed Fisher's test: "–" peaks of corresponding histone marks are better supported by RNA-Chrom contacts than "+" peaks

centage of RNAs with HiMoRNA peaks confirmed by at least one contact to 38.5%, 42.7%, 45.7%, 50.1%, and 53%. Some lncRNAs (e.g., MALAT1, HOXC-AS2, NEAT1, NR2F1-AS1, PVT1, MEG3) had nearly all HiMoRNA peaks confirmed with ±25 kb extension (*Fig. 3*). However, lncRNAs with a significantly lower proportion of peaks extended by ±25 kb and confirmed by contacts are more common (e.g., JPX, AP005263.1, MIR31HG) or have a proportion approaching 0 (e.g., MAPKAPK5-AS1). This discrepancy likely arises from the incomplete datasets of lncRNAs in HiMoRNA and RNA-Chrom, resulting from the stringent criteria used for prediction filtering and the limitations inherent in available experimental RNA-chromatin interaction data. For example, in RNA-Chrom, half of the lncRNAs considered in this study have fewer than 200 contacts (*Fig. 3B*), as most lncRNAs rely on "all-to-all" experimental data, which insufficiently captures contacts of lowly expressed RNAs.

HiMoRNA triads display either negative or positive correlations between lncRNA expression and the signal of epigenetic peaks ("–" and "+" peaks, respectively). To evaluate biological consistency with known data, we selected 30 lncRNAs and their corresponding histone peaks, for which "+" or "–" peaks of at least one histone mark are statistically significantly predominant (one-sided Fisher's exact test, p-value < 0.001), confirmed by RNA-Chrom contacts extended by ±25 kb (see "Experimental Section," subsection

"One-Sided Fisher's Exact Test," *Fig. 4*). After filtering the results by a p-value < 0.001, we obtained the following "lncRNA–histone mark" pairs.

Twenty-one lncRNAs with "+" peaks of corresponding histone marks were better supported by RNA-Chrom contacts than "–" peaks (right-tailed Fisher's exact test, p-value < 0.001).

Eleven lncRNAs with "–" peaks of corresponding histone marks were better supported by RNA-Chrom contacts than "+" peaks (left-tailed Fisher's exact test, p-value < 0.001).

Previous studies have shown the potential involvement of many identified lncRNAs in epigenetic regulation via histone modifications. Let us consider cases where "+" peaks are statistically significantly better supported by RNA-Chrom contacts than "–" peaks. For example, MIR4435-2HG is involved in establishing the activator mark H3K27ac in the enhancer region of the RPTOR locus [29]. Our data suggest that MIR4435-2HG, in addition to H3K27ac, likely targets other epigenetic modifications, such as H3K27me3, H3K36me3, H3K4me1, H3K4me2, H3K4me3, and H3K79me2 (*Fig. 4A*). Similarly, based on the data for MIR31HG [30], SNHG1, PVT1 [31–33], and the mouse lncRNA lnc-Nr2f1 (presumed to have functional conservation with human NR2F1-AS1) [34], we identified consistent histone modifications: NR2F1-AS1 – H3K27ac, MIR31HG – H3K4me3, SNHG1 – H3K27me3, PVT1 – H3K27me3. Additionally, we uncovered functional associations of these lncRNAs with other epigenetic marks: NR2F1-AS1 – H3K4me2, MIR31HG – H3K79me2, SNHG1 – H3K4me3, PVT1 – H3K4me1, H3K4me3, H3K9me3 (*Fig. 4A*).

Numerous H3K27me3 and H3K4me3 "+" peaks, validated through a chromatin interaction analysis, were identified for several lncRNAs (ZFAS1, SNHG4, SNHG1, SNHG3, PVT1, MIR4435-HG, and CYTOR). These peaks exhibit significantly stronger support from RNA-chromatin contacts than "–" peaks (*Fig. 4A*), representing opposing chromatin states. By analogy with known lncRNAs that establish both marks depending on their association with different effector proteins (e.g., ncRNA SRA [35], ANRIL [36]), it can be hypothesized that these lncRNAs also exhibit more complex mechanisms of chromatin activity regulation.

Cases where "–" peaks are statistically significantly better supported by RNA-Chrom contacts than "+" peaks can be explained by the corresponding lncRNAs regulating the removal of histone marks by recruiting demethylases and deacetylases to specific genomic loci (*Fig. 4B*). A lack of corroborating experimental data precludes a quality assessment of our predictions for these lncRNAs. We suggest that the "lncRNA–histone mark" pairs reported in this section (*Fig. 4*) are suitable candidates for further study.

Usage examples

The objective of integrating HiMoRNA and RNA-Chrom is to refine the functional relationship within the "lncRNA–epigenetic modification peak–associated gene" triads using data on the localization of the corresponding lncRNA in the genomic region near peaks of specific histone modifications. Below, we provide examples of user studies of several lncRNAs with known mechanisms of action.

lncRNA MIR31HG

The long non-coding RNA MIR31HG is a known regulator of the histone marks H3K4me1, H3K4me3, and H3K27ac. Previous studies have reported a reduction in the H3K4me1 and H3K27ac levels in the enhancer region of the *GLI2* gene and H3K4me3 and H3K27ac in the promoter region of the *FABP4* gene following MIR31HG knockdown [30, 37]. This observation can be validated using our integration of HiMoRNA and RNA-Chrom.

To this end, we created a query in HiMoRNA: lncRNA MIR31HG, histone marks H3K4me1 and H3K27ac, with the coordinates of the two selected genes specified with an extended promoter region of 10 kb in the genomic coordinates field (*Fig. 5A*). As a result, the HiMoRNA web resource generated a table with H3K27ac and H3K4me1 peaks correlated with MIR31HG expression across various tissues (*Fig. 5B*). We then selected a triad with an H3K27ac peak and navigated to the RNA-Chrom page displaying experimentally detected MIR31HG chromatin contacts in the region of the selected peak (by clicking "Go to RNA-Chrom DB," *Fig. 5C*). By selecting an RNA-chromatin experiment from the top table and clicking "All target genes" (*Fig. 6A*), we obtained a table that, in particular, reflected the interaction of MIR31HG with the *GLI2* gene (*Fig. 6B*). A step-by-step analysis is presented in *Supplementary Table 3*.

To explore the potential for novel biological insights into the lncRNA function, we hypothesized that integrating the HiMoRNA and RNA-Chrom datasets would reveal that MIR31HG regulates additional components of the Sonic hedgehog signaling pathway (KEGG:04340) besides *GLI2*. For this purpose, we used the KEGG Pathway database [38] to identify relevant genes. Next, we formulated a new HiMoRNA query consisting of lncRNA MIR31HG, histone modifications H3K4me1 and H3K27ac, and the 56 genes associated with the Hedgehog signaling pathway (*Supplementary Table 4*). The outcome was a table of 162 triads, which can be validated with the RNA-Chrom resource. For

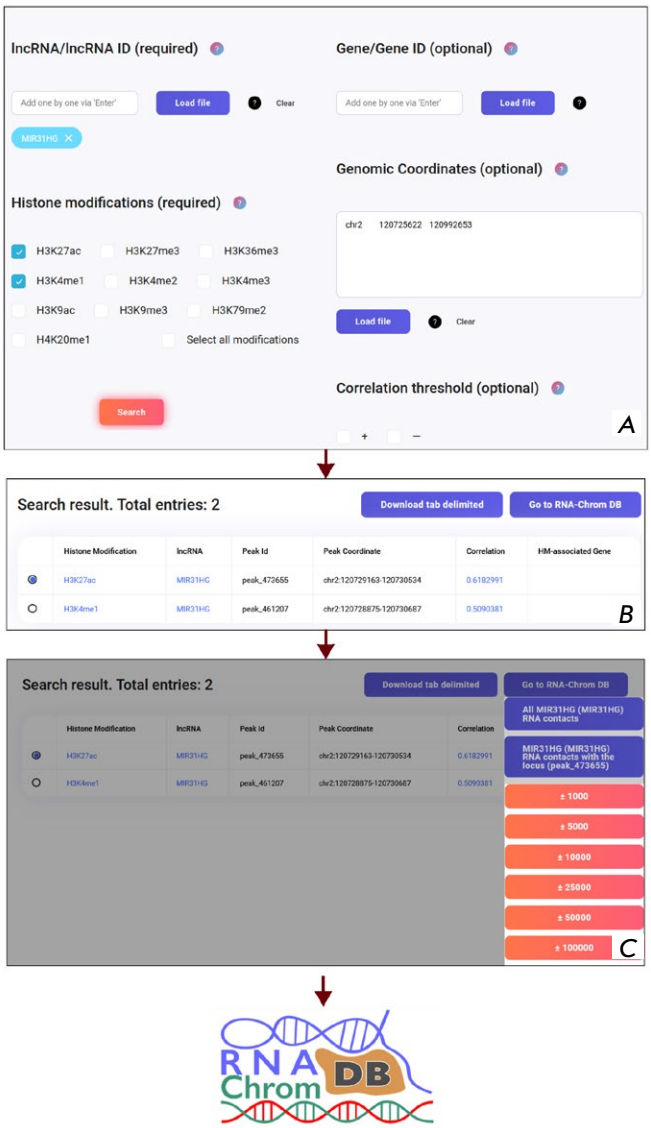


Fig. 5. Use case of the integration of the HiMoRNA and RNA-Chrom databases using the example of lncRNA MIR31HG. (*A*) Creating a query in HiMoRNA for MIR31HG, histone modifications H3K4me1 and H3K27ac, and genes *GLI2* and FABP4. (*B*) Table with search results. (*C*) Navigation to RNA-Chrom

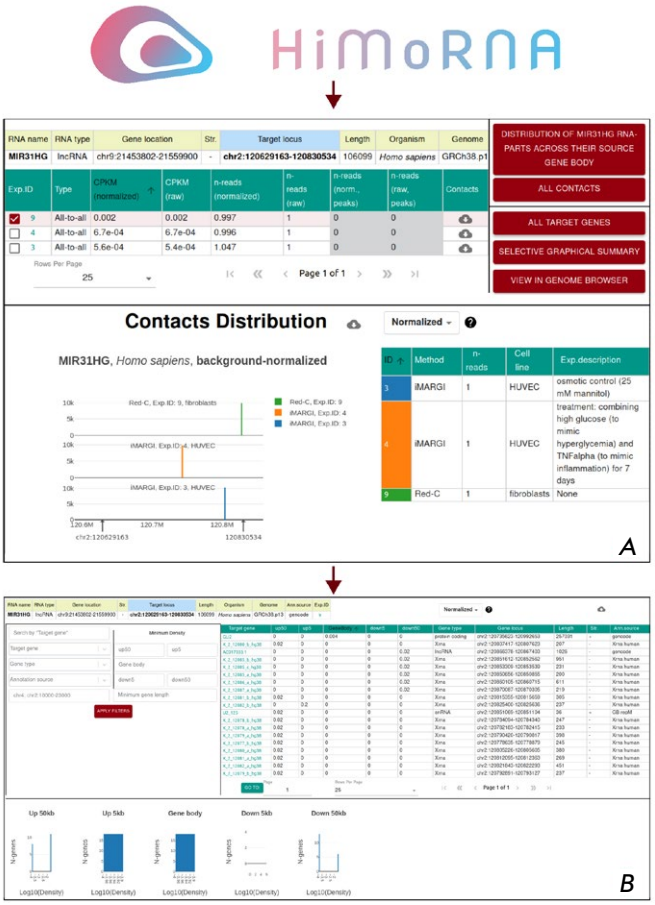


Fig. 6. Use case of the integration of the HiMoRNA and RNA-Chrom databases using the example of lncRNA MIR31HG. (*A*) RNA-Chrom page showing MIR31HG contacts with chromatin in the region of the extended HiMoRNA peak. (*B*) Table listing all genes in the region of the extended peak, indicating whether they interact with MIR31HG or not (Experiment ID: 9)

example, in the locus of the H3K27ac_963553 peak (chr9:95446174-95452554), MIR31HG interacts with the *PTCH1* gene, which encodes the "Sonic hedgehog" receptor. To determine if the gene set associated with the H3K27ac and H3K4me1 peaks, correlated with MIR31HG expression, exhibits significant enrichment of "Hedgehog signaling pathway" genes, a KEGG pathway enrichment analysis was conducted using the g:Profiler web resource [39]. The query included genes selected for MIR31HG and H3K27ac/H3K4me1, with all other genes associated with HiMoRNA peaks used as the background. The analysis revealed genes belonging to the "Hedgehog signaling pathway" to be enriched with H3K27ac peaks (p-value = 2.090×10^{-2}) but not with H3K4me1 peaks. This observation suggests the involvement of MIR31HG in regulating the "Hedgehog signaling pathway" through the establishment of the H3K27ac histone modification in the corresponding genomic loci.

lncRNA PVT1

The long non-coding RNA PVT1 is known to inhibit the expression of the *LATS2* gene in non-small cell lung cancer cells by recruiting EZH2 (a subu-

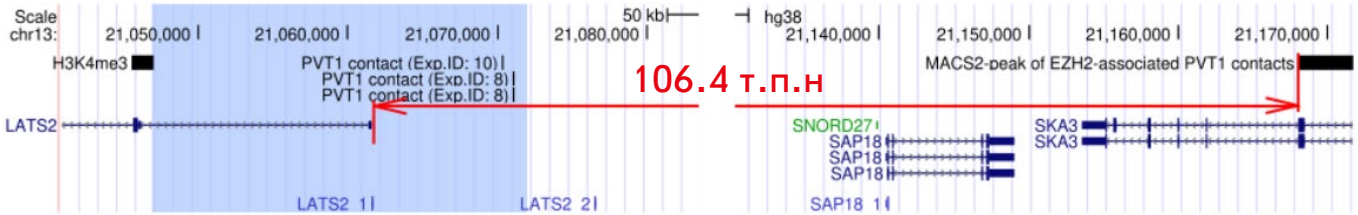


Fig. 7. Representation in the UCSC Genome Browser of the region encompassing the *LATS2* gene and its promoter vicinity, showing an H3K4me3 peak correlated with lncRNA PVT1 expression, lncRNA PVT1 contacts from two experiments (RNA-Chrom Exp. IDs: 8, 10), and an EZH2-mediated PVT1 contact peak. The blue region reflects the extension of the H3K4me3 peak coordinates by 25 kb, within which RNA-Chrom contacts were selected

nit of the PRC2 complex) to the corresponding promoter [40]. We performed a search for triads in HiMoRNA: lncRNA PVT1, all histone modifications, and the *LATS2* gene. Our findings revealed solely H3K4me3 activation peaks, exhibiting a negative correlation with PVT1 expression levels. This observation is consistent with previously published findings [39], wherein PVT1's recruitment of EZH2 contributes to the establishment of the repressive H3K27me3 mark. In RNA-Chrom, we observed contacts around one of the H3K4me3 peaks (peak_169403, chr13:21045571-21046978) in two experiments (K562 and MDA-MB-231 cell lines). Visualization of PVT1 contacts in the Genome Browser [41] confirms the presence of this peak in the promoter region of the *LATS2* gene (*Fig.* 7, step-by-step analysis presented in *Supplementary Table 5*). Additional confirmation of *LATS2* regulation by lncRNA PVT1 was obtained using Red-ChIP data (see "Experimental Section," subsection "Red-ChIP Data"). A peak of EZH2-mediated PVT1 contacts (chr13:21168000-21224000, q-value = 0.09) was identified 106.4 kb from the 5'-end of the *LATS2* gene (*Fig.* 7).

The absence of "lncRNA PVT1–H3K27me3 peak–*LATS2* gene" triads with a positive correlation in HiMoRNA is likely due to overly stringent filtering of H3K27me3 peaks during the database creation. The preceding examples from the "Use Cases" section illustrate the successful application of this integration in generating testable hypotheses regarding lncRNA involvement in the epigenetic regulation of specific genes.

DISCUSSION

Prospects for Use and Limitations of the Approach

The HiMoRNA database comprises genomic loci exhibiting a significant correlation between histone modification signals and lncRNA expression across diverse cell types and tissues. Currently, it contains over 5 million correlations for 10 types of histone modifications and 4,145 lncRNAs. We hypothesize that some of these correlations may represent false positives or indirect regulatory relationships, thus requiring further validation using external data. The RNA-Chrom database houses genome-wide RNA-chromatin interaction data. These data unfortunately lack sufficient representation of contacts from lowly expressed ncRNAs and are heavily biased towards contacts from nascent transcripts. Furthermore, these data are insufficient to formulate hypotheses regarding the functional roles of these interactions. With these limitations in mind, integrating the data from HiMoRNA and RNA-Chrom is a reasonable approach to characterize the impact of lncRNAs on epigenetic modifications and gene expression.

Despite the advantages of integration, some inconsistencies remain. No H3K27ac or H3K4me1 peaks were found in the *FABP4* gene locus in the HiMoRNA database. This observation contradicts experimental data and the RNA-Chrom database. There are other negative examples in both HiMoRNA and RNA-Chrom for well-known lncRNAs involved in epigenetic regulation and chromatin structure maintenance. For instance, no H3K27me3 peaks correlating with MEG3 were observed in HiMoRNA, despite evidence suggesting that MEG3 regulates the PRC2 complex and contributes to maintaining H3K27me3, particularly in the promoter regions of the *SMAD2*, *TGFB2*, and *TGFBR1* genes [11]. The absence of such peaks is likely attributable to the cell-specific expression of most lncRNAs, with the mentioned mechanism observed in a cell type not represented in HiMoRNA. Even when ChIP-seq data are available, the standard peak-calling procedure may be too stringent, potentially filtering out biologically significant interactions.

We observed instances where HiMoRNA predictions aligned with published experimental data, but the corresponding RNA-chromatin contacts were absent from the relevant genomic locus in the RNA-Chrom dataset. For example, for lncRNA MAPKAPK5-AS1, most correlated peaks from HiMoRNA are not supported by the expression of this RNA, resulting in a low number of observed chromatin contacts in "all-to-all" experiments. We posit that these instances may stem from variations in cell types across the two databases as a result of insufficient data.

Due to insufficient experimental data, neither database contains exhaustive information. Therefore, some documented biological examples might have been overlooked in the integration process. Nevertheless, their integration offers complementary strengths, such as mitigating various systematic errors that are due to the multi-omics nature of the combined data and expanding the generation of interpretable hypotheses about the mechanisms of epigenetic regulation of gene expression by long non-coding RNAs.

FURTHER DEVELOPMENT

The implemented integration could be significantly improved through the incorporation of supplementary genome-wide data and annotations. The dataset may contain information regarding the three-dimensional chromatin structure, gene expression and co-expression patterns (including long non-coding RNA expression, the target genes of the triad, and the genes associated with histone modification), in addition to the localization of DNA-binding and chromatin-modifying proteins. Given the current scarcity of such experimental data, it is worth considering the use of bioinformatics predictions. One potential direction could involve incorporating results from predictions of the type of lncRNA-chromatin interactions (for a comparison of programs determining the mechanisms of lncRNA interactions with other molecules; see, for example, [42]). The combination of predicted lncRNA–target interactions and multi-omics experimental data has facilitated effective hypothesis generation concerning the roles of specific lncRNAs [43–45]. Another important direction would be to include data on gene expression changes following artificial alterations in the concentration of specific lncRNAs in cells [8], as well as experimentally validated information on the involvement of specific lncRNAs in regulating particular histone modifications [46]. This would provide an additional layer of validation for the results of the HiMoRNA and RNA-Chrom integration. Furthermore, from a practical perspective, it would be useful to enhance the integration with an assessment of the statistical significance of the co-localization of HiMoRNA peaks and RNA-Chrom contacts for a specific lncRNA using specialized software tools such as Genometricorr [47], StereoGene [48], and RegioneR [49].

The field of lncRNA research is rapidly evolving. We will maintain support for the integration of HiMoRNA and RNA-Chrom as both databases expand their taxonomic scope and incorporate updated data. Upon the emergence of new, experimentally validated data, we intend to construct multiple predictive models for "lncRNA–histone epigenetic modifications–associated gene" interactions. We are confident that the continued collaborative expansion of the HiMoRNA and RNA-Chrom web resources will contribute to a deeper understanding of the functional role of lncRNAs in the epigenetic regulation of genes. ●

The authors express their gratitude to A. Nikolskaya for providing processed Red-ChIP data and to anonymous reviewers for their valuable suggestions.

This work was supported by the Russian Science Foundation grant No. 23-14-00371 (PI: Yu.A. Medvedeva).

Author Contributions

Improvement of the HiMoRNA web interface and database, software development, I.I.; adaptation of the RNA-Chrom web resource for integration with HiMoRNA, gene name correspondence, project administration, G.R.; use cases, G.R. and D.M.; general supervision, A.M. and Yu.M.; manuscript preparation, I.I., G.R., D.M., A.M., and Yu.M. All authors have read and agreed to the published version of the manuscript.

Supplementaries are available on the website https://doi.org/10.32607/actanaturae.27543.

REFERENCES

1. Carninci P., Sandelin A., Lenhard B., Katayama S., Shimokawa K., Ponjavic J., Semple C.A.M., Salmena L., Nishida M., Hayashizaki Y., et al. // Science. 2005. V. 309. № 5740. P. 1559–1563. doi: 10.1126/science.1112014.
2. Hon C.-C., Ramilowski J.A., Harshbarger J., Bertin N., Rackham O.J.L., Garmire L.X., Forrest A.R.R., Carninci P., Kawaji H., Hayashizaki Y., et al. // Nature. 2017. V. 543. № 7644. P. 199–204. doi: 10.1038/nature21374.
3. Cabili M.N., Trapnell C., Goff L., Koziol M., Tazon-Vega

B., Regev A., Rinn J.L. // Genes Dev. 2011. V. 25. № 18. P. 1915–1927. doi: 10.1101/gad.17446611.
4. Andersson R., Gebhard C., Miguel-Escalada I., Hoof I., Bornholdt J., Boyd M., Chen Y., Zhao X., Schmidl C., Suzuki T., et al. // Nat. Commun. 2014. V. 5. № 1. P. 5336. doi: 10.1038/ncomms6336.
5. Ulitsky I., Shkumatava A., Jan C.H., Sive H., Bartel D.P. // Cell. 2011. V. 147. № 7. P. 1537–1550. doi: 10.1016/j.cell.2011.11.055.
6. Quinn J.J., Ilik I.A., Qu K., Georgiev P., Chu C., Akhtar A., Chang H.Y. // Genes Dev. 2016. V. 30. № 2. P. 191–207. doi: 10.1101/gad.272187.115.
7. Alam T., Medvedeva Y.A., Jia H., Brown J.B., Lipovich L., Baillie J.K. // PLoS One. 2014. V. 9. № 10. P. e109443. doi: 10.1371/journal.pone.0109443.
8. Ramilowski J.A., Yip C.W., Agrawal S., Chang J.-C., Ciani Y., Kulakovskiy I.V., Mendez M., Ooi J.L.C., Ouyang J.F., Nguyen A., et al. // Genome Res. 2020. V. 30. № 7. P. 1060–1072. doi: 10.1101/gr.254516.119.
9. Khalil A.M., Guttman M., Huarte M., Garber M., Raj A., Rivea Morales D., Thomas K., Presser A., Bernstein B.E., van Oudenaarden A., et al. // Proc. Natl. Acad. Sci. USA. 2009. V. 106. № 28. P. 11667–11672. doi: 10.1073/pnas.0904715106.
10. Grote P., Herrmann B.G. // RNA Biol. 2013. V. 10. № 10. P. 1579–1585. doi: 10.4161/rna.26165.
11. Mondal T., Subhash S., Vaid R., Enroth S., Uday S., Reinius B., Mitra S., Mohammed A., James A.R., Hoberg E., et al. // Nat. Commun. 2015. V. 6. № 1. P. 7743. doi: 10.1038/ncomms8743.
12. Goff L.A., Rinn J.L. // Genome Res. 2015. V. 25. № 10. P. 1456–1465. doi: 10.1101/gr.191122.115.
13. Mazurov E., Sizykh A., Medvedeva Y.A. // Non-Coding RNA. 2022. V. 8. № 1. P. 18. doi: 10.3390/ncrna8010018.
14. Engreitz J.M., Pandya-Jones A., McDonel P., Shishkin A., Sirokman K., Surka C., Kadri S., Xing J., Goren A., Lander E.S., et al. // Science. 2013. V. 341. № 6147. P. 1237973. doi: 10.1126/science.1237973.
15. Simon M.D., Wang C.I., Kharchenko P.V., West J.A., Chapman B.A., Alekseyenko A.A., Borowsky M.L., Kuroda M.I., Kingston R.E. // Proc. Natl. Acad. Sci. USA. 2011. V. 108. № 51. P. 20497–20502. doi: 10.1073/pnas.1113536108.
16. Chu C., Qu K., Zhong F.L., Artandi S.E., Chang H.Y. // Mol. Cell. 2011. V. 44. № 4. P. 667–678. doi: 10.1016/j.molcel.2011.08.027.
17. Quinn J.J., Qu K., Chang H.Y. // Nat. Biotechnol. 2014. V. 32. № 9. P. 933–940. doi: 10.1038/nbt.2943.
18. Chu H.-P., Cifuentes-Rojas C., Kesner B., Aeby E., Lee H.- G., Wei C., Oh H.J., Boukhali M., Haas W., Lee J.T., et al. // Cell. 2017. V. 170. № 1. P. 86–101. doi: 10.1016/j.cell.2017.06.017.
19. Sridhar B., Rivas-Astroza M., Nguyen T.C., Chen W., Yan Z., Cao X., Hebert L., Zhong S. // Curr. Biol. 2017. V. 27. № 4. P. 602–609. doi: 10.1016/j.cub.2017.01.011.
20. Li X., Zhou B., Chen L., Gou L.-T., Li H., Fu X.-D. // Nat. Biotechnol. 2017. V. 35. № 10. P. 940–950. doi: 10.1038/nbt.3968.
21. Bell J.C., Jukam D., Teran N.A., Risca V.I., Smith O.K., Johnson W.L., Skotheim J.M., Greenleaf W.J., Straight A.F. // eLife. 2018. V. 7. P. e27024. doi: 10.7554/eLife.27024.
22. Yan Z., Huang N., Wu W., Chen W., Jiang Y., Zhang J., Zhang X., Zhang Y., Zhang Q., Zhang L., et al. // Proc. Natl. Acad. Sci. USA. 2019. V. 116. № 8. P. 3328–3337. doi: 10.1073/pnas.1819788116.
23. Bonetti A., Agostini F., Suzuki A.M., Hashimoto K., Pascarella G., Gimenez J., Roos L., Nash A.J., Ghazanfar S., Carninci P., et al. // Nat. Commun. 2020. V. 11. № 1. P. 1018. doi: 10.1038/s41467-020-14337-6.
24. Gavrilov A.A., Zharikova A.A., Galitsyna A.A., Luzhin A.V., Rubanova N.M., Golov A.K., Petrova N.V., Kantidze O.L., Ulianov S.V., Misteli T., et al. // Nucl. Acids Res. 2020. V. 48. № 12. P. 6699–6714. doi: 10.1093/nar/gkaa456.
25. Ryabykh G.K., Zharikova A.A., Galitsyna A.A., Ulianov S.V., Razin S.V., Gavrilov A.A. // Mol. Biol. 2022. V. 56. № 2. P. 210–228. doi: 10.1134/S002689332202017X.
26. Ryabykh G.K., Dikstein N., Zharikova A.A., Galitsyna A.A., Ulianov S.V., Razin S.V., Gavrilov A.A. // Database J. Biol. Databases Curation. 2023. V. 2023. P. baad025. doi: 10.1093/database/baad025.
27. Gavrilov A.A., Golov A.K., Luzhin A.V., Zharikova A.A., Galitsyna A.A., Rubanova N.M., Kantidze O.L., Ulianov S.V., Razin S.V., Misteli T., et al. // Proc. Natl. Acad. Sci. USA. 2022. V. 119. № 1. P. e2116222119. doi: 10.1073/pnas.2116222119.
28. Mylarshchikov D.E., Baumgart S.J., Mazurov E., Saveliev A., Medvedeva Y.A. // NAR Genomics Bioinforma. 2024. V. 6. № 2. P. lqae054. doi: 10.1093/nargab/lqae054.
29. Hartana C.A., Rassadkina Y., Gao C., Yashiro-Ohtani Y., Mercier F., Deng S., Li X., Lin 7. R.H., Das S., Lian C., et al. // J. Clin. Invest. 2021. V. 131. № 9. P. e146136. doi: 10.1172/JCI146136.
30. Chen W., Zhang J., Xu H., Dai X., Zhang X., Wang W., Wu Y., Li Y., Wang X., Zhang Y., et al. // Oncogene. 2024. V. 43. № 18. P. 1328–1340. doi: 10.1038/s41388-024-02986-2.
31. Li B., Jiang Q., Liu X., Yang X., Li Z., Li J., Zhang Y., Wu Y., Li X., Zhang Y., et al. // Cell Death Dis. 2020. V. 11. № 10. P. 823. doi: 10.1038/s41419-020-03031-6.
32. Li Z., Guo X., Wu S. // Stem Cell Res. Ther. 2020. V. 11. № 1. P. 435. doi: 10.1186/s13287-020-01953-8.
33. Nylund P., Gimenez G., Laine I., Massinen S., Gordon S., Kuusela M., Kontturi S., Rusanen A., Vettenranta K., Lohi O., et al. // Haematologica. 2024. V. 109. № 2. P. 567–577. doi: 10.3324/haematol.2022.282375.
34. Ang C.E., Trevino A.E., Chang S., Soeung M., Ma L., Chidambaram S., Young C., Wernig M., Sudhof T.C., Chang H.Y., et al. // eLife. 2019. V. 8. P. e41770. doi: 10.7554/eLife.41770.
35. Wongtrakoongate P., Riddick G., Hashem O., Harris R., Jones S., Ramjaun A., Bunjobpol W., Felsenfeld G., Turner B.M., Schwabe J.W.R., et al. // PLoS Genet. 2015. V. 11. № 10. P. e1005615. doi: 10.1371/journal.pgen.1005615.
36. Alfeghaly C., Bozec D., Forterre M., Muller Q., Bernard D., Munch M., Schulte M.L., Nollet M., Paul N., Regnault B., et al. // Nucl. Acids Res. 2021. V. 49. № 9. P. 4954–4970. doi: 10.1093/nar/gkab267.
37. Huang Y., Zheng Y., Jia L., Li W. // Sci. Rep. 2017. V. 7. № 1. P. 8080. doi: 10.1038/s41598-017-08131-6.
38. Kanehisa M., Furumichi M., Tanabe M., Sato Y., Morishima K. // Nucl. Acids Res. 2016. V. 44. № D1. P. D457–D462. doi: 10.1093/nar/gkv1070.
39. Kolberg L., Raudvere U., Kuzmin I., Vilo J., Peterson H. // Nucl. Acids Res. 2023. V. 51. № W1. P. W207–W212. doi: 10.1093/nar/gkad347.
40. Wan L., Zhang L., Fan K., Cheng Z.X., Sun Q.C., Wang J.J. // Mol. Cancer Ther. 2016. V. 15. № 5. P. 1082–1094. doi: 10.1158/1535-7163.MCT-15-0509.
41. Kent W.J., Sugnet C.W., Furey T.S., Roskin K.M., Pringle T.H., Zahler A.M., Haussler D. // Genome Res. 2002. V. 12. № 6. P. 996–1006. doi: 10.1101/gr.229102.
42. Antonov I.V., Mazurov E., Borodovsky M., Medvedeva

Y.A. // Brief. Bioinform. 2019. V. 20. № 2. P. 551–564. doi: 10.1093/bib/bbx132.
43. Antonov I., Medvedeva Y.A. // Genes. 2020. V. 11. № 12. P. E1483. doi: 10.3390/genes11121483.
44. Matveishina E., Antonov I., Medvedeva Y.A. // Int. J. Mol. Sci. 2020. V. 21. № 3. P. 830. doi: 10.3390/ijms21030830.
45. Ogunleye A.J., Romanova E., Medvedeva Y.A. // F1000Research. 2021. V. 10. P. 204. doi: 10.12688/f1000research.51844.1.
46. Marakulina D., Vorontsova Y., Mazurov E., Anisenko A., Korchagina A., Medvedeva Y.A., Kolmykov S., Fishman V., Kulakovskiy I.V., Meshcheryakov R., et al. // Nucl. Acids Res. 2023. V. 51. № D1. P. D564–D570. doi: 10.1093/nar/gkac809.
47. Favorov A., Mularoni L., Cope L.M., Medvedeva Y.A., Mironov A.A., Makeev V.J., Wheelan S.J. // PLoS Comput. Biol. 2012. V. 8. № 5. P. e1002529. doi: 10.1371/journal.pcbi.1002529.
48. Stavrovskaya E.D., Tepliuk N., Mironov A.A., Panchenko A.R., Favorov A.V., Makeev V.J. // Bioinforma. Oxf. Engl. 2017. V. 33. № 20. P. 3158–3165. doi: 10.1093/bioinformatics/btx379.
49. Gel B., Díez-Villanueva A., Serra E., Buschbeck M., Peinado M.A., Malinverni R. // Bioinforma. Oxf. Engl. 2016. V. 32. № 2. P. 289–291. doi: 10.1093/bioinformatics/btv562.

Monomeric α-Synuclein Real-Time Induced Conversion: A New Approach to the Diagnostics of Neurodegenerative Synucleinopathies with Weak RT-QuIC Responses

D. A. Orlova[1*], A. A. Kudriaeva[1], N. A. Kolotyeva[2], E. O. Ivanova[2], E. Yu. Fedotova[2], P. P. Tregub[2,3], A. B. Salmina[2], S. N. Illarioshkin[2], A. A. Belogurov Jr.[1,4]

[1]Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, 117997 Russia

[2]Brain Science Institute, Research Center of Neurology, Moscow, 125367 Russia

[3]Department of Pathophysiology, Sechenov First Moscow State Medical University, Moscow, 119991 Russia

[4]Department of Biological Chemistry, Russian University of Medicine, Moscow, 127473 Russia

*E-mail: dorlova01@yandex.ru

Received: October 06, 2024; in final form, February 07, 2025

DOI: 10.32607/actanaturae.27530

ABSTRACT **Neurodegenerative disorders classified as synucleinopathies (Parkinson's disease, dementia with Lewy bodies, and multiple-system atrophy) are characterized by the accumulation of aberrant α-synuclein aggregates in neurons and glial cells. These diseases manifest clinically several years after the initial formation of pathological protein aggregates in the brain, making early and accurate diagnosis challenging. In recent years, a new method, which is based on real-time quaking-induced conversion (RT-QuIC) of α-synuclein, has been developed and validated. This technology holds great promise as a powerful diagnostic tool for the early and precise identification of synucleinopathies, potentially opening new horizons in the study of neurodegenerative diseases. RT-QuIC detects misfolded α-synuclein aggregates in human physiological fluids by introducing an excess of recombinant α-synuclein, which undergoes conformational conversion in an exponential, prion-like manner. The production of high-quality recombinant α-synuclein is a critical step in the effective application of this method, as protein purity significantly affects the sensitivity and specificity of the assay — key factors in its diagnostic utility. Using a three-step chromatographic purification protocol, we produced recombinant monomeric α-synuclein with a purity exceeding 97% from the periplasmic fraction of bacterial cells. While higher purity increases the assay duration, it also reduces the background signal and permits extended incubation times, which are essential for reliably detecting synucleinopathies with weak RT-QuIC responses, such as the cerebellar subtype of multiple-system atrophy. The data presented support the conclusion that optimized components of the RT-QuIC system will enable an accurate diagnosis of neurodegenerative synucleinopathies.**

KEYWORDS **α-synuclein, synucleinopathies, multiple system atrophy, Lewy body dementia, real-time quaking-induced conversion (RT-QuIC), diagnostics.**

ABBREVIATIONS **CSF – cerebrospinal fluid; LBD – Lewy body dementia; EDTA – ethylenediaminetetraacetic acid; FPLC – fast protein liquid chromatography; SEC – size exclusion chromatography; HIC – hydrophobic interaction chromatography; IEC – ion exchange chromatography; IPTG – isopropyl-β-D-1-thiogalactopyranoside; MRI – magnetic resonance imaging; MSA–C – multiple system atrophy of the cerebellar type; MSA – multiple system atrophy; MSA–P – multiple system atrophy of the parkinsonian type; PD – Parkinson's disease; PAGE – polyacrylamide gel electrophoresis; RT-QuIC – real-time quaking-induced conversion; SAA – seed amplification assay; SPS – stiff-person syndrome; ThT – thioflavin T.**

INTRODUCTION

Synucleinopathies are a group of neurodegenerative diseases that include Parkinson's disease, Lewy body dementia (LDB), and multiple system atrophy (MSA). The aggregation of a misfolded α-synuclein protein in neurons and/or glial cells plays a key role in the pathogenesis of these diseases: α-synuclein with an aberrant conformation has been found to be capable of trans-synaptic spreading throughout the central nervous system, like prions [1–4]. α-Synuclein is a 14 kDa presynaptic protein encoded by the *SNCA* gene located on the long arm of chromosome 4 at locus 4q21–22. α-Synuclein is predominantly expressed in the midbrain substantia nigra, neocortex, and hippocampus [5]. Physiological α-synuclein levels are essential for normal mitochondrial functioning, neurotransmitter release, and maintenance of morphological cell integrity. Overexpression of α-synuclein and changes in its aggregation properties result in mitochondrial dysfunction, neuroinflammation, and impaired synaptic release of dopamine and other neurotransmitters, leading to neuronal death [6, 7]. A distinctive feature of the members of the synuclein family is their tendency to form aggregates. Native α-synuclein is an unstructured, monomeric soluble protein. In pathological conditions, it forms β-pleated oligomers (protofibrils) that are subsequently transformed into amyloid fibrils and deposited in neurons in the form of Lewy bodies and neurites, as well as other inclusions [8–10]. The mechanism of α-synuclein aggregate growth in each case is thought to be based on seed polymerization. Trans-synaptic spread of aberrant molecules from neuron to neuron is observed in Parkinson's disease and LBD, whereas their accumulation and transmission in glial cells occurs in MSA [11].

To date, there has been no standardized reference method for detecting α-synuclein aggregates in the nervous system. Existing immunohistochemical approaches for identifying α-synuclein in peripheral tissue biopsies (e.g., skin, salivary glands, etc.) are technically complex and prohibitively expensive [12], which limits their routine use in clinical practice. Meanwhile, the development and implementation of highly sensitive techniques for detecting pathological forms of α-synuclein and other brain-derived proteins are critically important for improving diagnostic accuracy, particularly at the prodromal stage and for enabling timely therapeutic interventions in neurodegenerative diseases. One promising approach is the seed amplification assay (SAA), originally developed for the prion disease. This method exploits a protein misfolding chain reaction triggered by the presence of pathological protein conformers in patient-derived biological samples added to a reaction medium. Applying the prion hypothesis to α-synuclein in Parkinson's disease, MSA, and DLB has spurred research into the potential of SAA to detect pathological α-synuclein conformers in various tissues and body fluids, including the skin, olfactory mucosa, cerebrospinal fluid, and the blood [12, 13].

A modern SAA version is the real-time quaking-induced conversion (RT-QuIC) assay [14, 15]. This assay uses a recombinant protein as a substrate, and the patient's biological material serves as a "seed" to detect protein misfolding. RT-QuIC is based on the ability of a pathological α-synuclein form to induce conformational changes in normal monomeric α-synuclein, which leads to misfolded protein aggregation. The assay principle is to create artificial conditions for the seed amplification of α-synuclein by alternating the incubation and quaking cycles, which promotes additional fragmentation of the formed aggregates and an increase in protofibril formation. α-Synuclein aggregation is detected using a fluorescent dye, thioflavin T (ThT), that is incorporated into the aggregates during polymerization, which increases fluorescence over time [14, 16]. It should be noted that the purity of the substrate, monomeric soluble α-synuclein, is crucial for the reliability and reproducibility of results, as well as for the prevention of false-positive reactions [17, 18].

This study was aimed at improving the RT-QuIC-based system for the diagnostics of neurodegenerative synucleinopathies and, in particular, at developing a method for the production of highly purified recombinant monomeric wild-type α-synuclein for its further use as a substrate in the RT-QuIC assay.

EXPERIMENTAL

Expression of recombinant α-synuclein in *Escherichia coli* cells

The pET33b+ plasmid containing the human α-synuclein gene sequence was transformed into One Shot BL21 (DE3) Star *E. coli* cells (Thermo Fisher Scientific, USA). The cells were cultured in 500 mL of a bacterial LB medium containing 50 μg/mL kanamycin and 0.1% glucose under constant stirring at 200 rpm. The culture was grown to an optical density of 0.6 at a wavelength of 600 nm (OD_{600} = 0.6). Expression of the target protein was induced by adding isopropyl-β-D-l-thiogalactopyranoside (IPTG) to a final concentration of 1 mM, followed by incubation of the cells at 37°C and vigorous stirring for 4 h. Following the expression, the cells were pelleted by centrifugation at 4,000 *g* for 15 min.

Periplasmic lysis

After centrifugation, the cell pellet produced from 300 mL of the culture medium was resuspended in 60 mL of an osmotic shock buffer (30 mM Tris, 40% sucrose, and 2 mM EDTA, pH 7.2) and incubated at room temperature for 10 min. The suspension was centrifuged at 18,000 *g* for 20 min, the supernatant was separated, and the pellet was resuspended in 50 mL of ice water (dH_2O) containing 20 μL of a saturated $MgCl_2$ solution. The resulting suspension was kept on ice for 3 min, followed by centrifugation at 18,000 *g* for 20 min. The supernatant was dialyzed against a buffer containing 10 mM Tris and 1 mM EDTA (pH 7.2) at 4°C overnight.

Ion exchange chromatography

Ion exchange chromatography (IEC) was performed using a C 10/10 column (Cytiva, USA) packed with the Q Sepharose Fast Flow sorbent (Cytiva, USA) by means of a BioLab 30 fast protein liquid chromatography (FPLC) system (Jiangsu Hanbon Science and Technology Co., Ltd, China). Before loading the protein sample, the column was equilibrated with an IEC A buffer (10 mM Tris, pH 7.2). Prior to chromatography, all buffer solutions and protein samples were degassed and filtered through a 0.22 μm membrane filter. Elution was performed using a linear gradient of IEC A (10 mM Tris, pH 7.2) and IEC B (10 mM Tris and 0.15 M $(NH_4)_2SO_4$, pH 7.2) buffers (7 column volumes), followed by a final column wash with a 100% IEC B buffer. The optical density of the eluate was monitored at a wavelength of 280 nm. To determine the time of α-synuclein desorption from the chromatographic column, the resulting fractions were collected and resolved on a 13% polyacrylamide gel electrophoresis (PAGE) under denaturing conditions, followed by staining with the Coomassie brilliant blue. Fractions containing protein bands corresponding to the molecular weight of monomeric α-synuclein were pooled and dialyzed against a 20 mM Tris buffer, pH 7.0 and 0.15$(NH_4)_2SO_4$, overnight.

Hydrophobic interaction chromatography

Hydrophobic interaction chromatography (HIC) was performed using a C 10/10 column (Cytiva) packed with the Phenyl Sepharose High Performance sorbent (Cytiva). Before loading the protein sample, the column was equilibrated with HIC buffer A (50 mM bis-Tris and 1 M $(NH_4)_2SO_4$, pH 7.0). The salt concentration in the samples was adjusted to 1 M by gradually adding (NH4)2SO4 while stirring the mixture at 4°C; the pH of the sample was adjusted to 7.0. Then, a sample was loaded onto the chromatographic column and eluted using a linear gradient of the HIC A (50 mM bis-Tris and 1 M $(NH_4)_2SO_4$, pH 7.0) and HIC B (50 mM bis-Tris, pH 7.0) buffers (7 column volumes), followed by a final column wash with a 100% HIC B buffer. α-Synuclein-containing fractions were pooled and dialyzed against a 20 mM Tris buffer (pH 7.2) at 4°C overnight. The resulting protein solution was concentrated to 1.0–1.5 mg/mL using centrifugal concentrators with a cutoff of 5,000 Da and frozen at –80°C until further experiments.

Gel-filtration chromatography

Gel-filtration chromatography (GFC) was performed using a Superose 12 10/30 FPLC column (GE Pharmacia, USA). Before loading the protein, the column was equilibrated with a buffer (20 mM Tris, pH 7.2). Some 500 μL of a pre-concentrated sample was loaded onto the column at a flow rate of 1 mL/min. Chromatographic fractions were resolved on 13% PAGE under denaturing conditions, followed by staining with the Coomassie brilliant blue. The gel image was analyzed using the Image Lab Touch software for densitometric determination of protein purity. The resulting recombinant α-synuclein was aliquoted to the desired volume and stored at –80°C until the RT-QuIC assay.

Sample collection and preparation

To conduct pilot studies using the RT-QuIC technology, cerebrospinal fluid (CSF) samples (n = 3) were collected from patients with MSA (cerebellar type), LBD, and stiff-person syndrome (control), aged 58–69 years. Diagnoses were made based on anamnesis data, clinical examination, and the results of special laboratory and instrument tests, including high-field MRI (3 T) in the appropriate research modes. All patients gave written consent for the examination. The study was approved by the local ethics committee of the Research Center of Neurology (protocol No. 7-1/24). Lumbar puncture was performed in the morning, after overnight fasting. CSF was sterile-collected into a polypropylene tube, centrifuged, aliquoted to 500 μL portions, flash frozen, and stored at –80°C.

RT-QuIC

RT-QuIC assays were performed in black 96-well plates with an opaque bottom. Each well contained 37 ± 3 mg of glass beads (600–800 μm), a reaction buffer (100 mM phosphate buffer, pH 8.2, 10 μM ThT) containing recombinant α-synuclein at a final concentration of 0.1 mg/mL, and an undiluted CSF sample. The plate was sealed with an adhesive tape and placed in a ClarioStar multimodal plate reader (BMG Labtech). Samples were incubated at 37°C with intermittent shaking cycles for 125 h. Fibril forma-

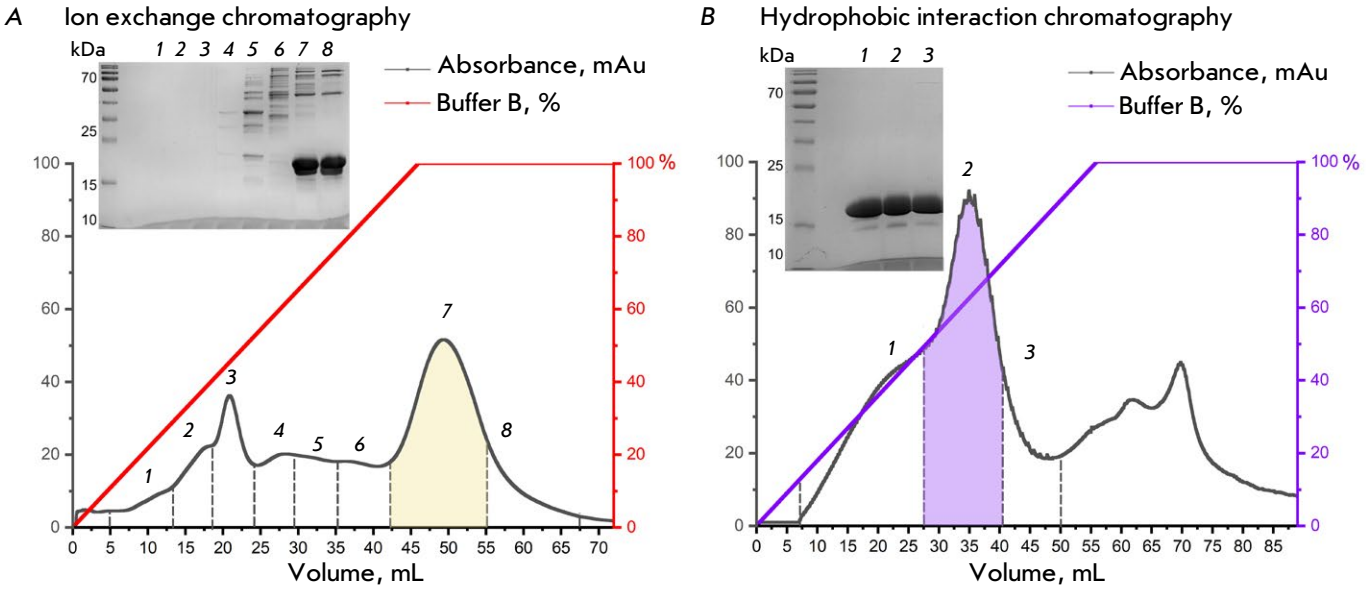


Fig. 1. (*A*) Ion exchange chromatography of samples after periplasmic lysis. The peak highlighted in color corresponds to the fraction containing the largest amount of α-synuclein. The inset at the top left-hand side shows the electrophoresis of protein fractions after ion exchange chromatography. (*B*) Hydrophobic interaction chromatography of samples after ion exchange chromatography. The peak highlighted in color corresponds to the fraction containing the highest amount of α-synuclein. The inset at the top left-hand side shows the electrophoresis of protein fractions after hydrophobic interaction chromatography. The molecular weight of the target protein is 19 kDa (apparent molecular weight). The peak number on the chromatographic profiles corresponds to the PAGE lane number

tion kinetics were monitored by measuring the ThT fluorescence intensity at 450/480 nm every 60 min. Measurements were discontinued when the ThT fluorescence signal reached a plateau. Each sample was analyzed in triplicate.

RESULTS

The first step in the chromatographic purification of recombinant α-synuclein was ion exchange chromatography. The purity and separation efficiency of the protein fractions obtained during chromatography were assessed by electrophoretic analysis in a polyacrylamide gel in the presence of sodium dodecyl sulfate, followed by staining the gel with the Coomassie brilliant blue. Analysis of the fractions revealed the elution profile of α-synuclein from the chromatographic column. The main fraction containing α-synuclein is depicted in color on the chromatogram (*Fig. 1A*). According to electrophoretic analysis (*Fig. 1A*, inset), protein elution started at an IEC B buffer concentration of 60% in the eluent (with correction for the column volume). A further increase in the concentration of the IEC B buffer to 80% led to complete elution of α-synuclein. The α-Synuclein-containing fractions 7 and 8 were pooled and used in further steps in the protein purification.

α-Synuclein isolated from the periplasm contained protein impurities, so an additional purification step was required to prepare a homogeneous product. For this purpose, a hydrophobic chromatography step was introduced. Analysis of the chromatographic elution profile of α-synuclein showed that desorption of the target protein from hydrophobic sorbent started at an HIC B buffer concentration of 15% in the mobile phase and continued up to 65%. The fraction containing the largest amount of α-synuclein is depicted in color on the chromatogram (*Fig. 1B*). The efficiency of protein purification by hydrophobic chromatography

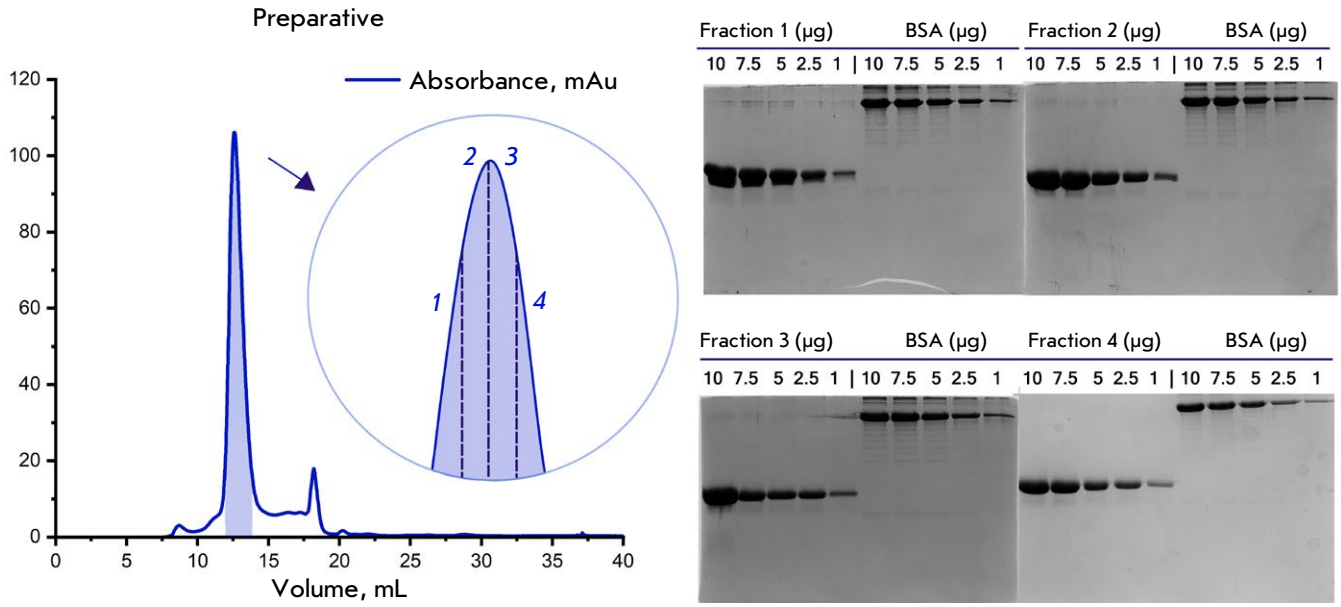


Fig. 2. Size exclusion chromatography of α-synuclein. The peak highlighted in color corresponds to the fraction containing the target protein. At the top right-hand side, images of the denaturing electrophoresis of protein fractions are shown. α-Synuclein fractions at different concentrations (10, 7.5, 5, 2.5, and 1 μg) are presented on the left, and the molecular weight marker (BSA) at the same concentrations is presented on the right. The fraction numbers on the chromatogram correspond to those on PAGE images

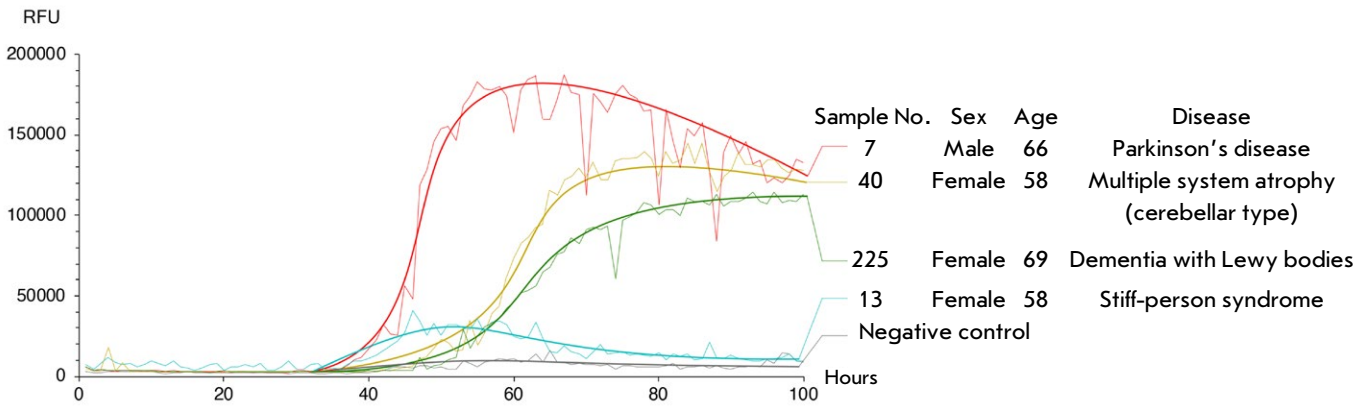


Fig. 3. Curves of pathological α-synuclein amplification in patients (RFU – relative fluorescence units). Analysis of samples from patients with stiff-person syndrome (blue), Lewy body dementia (green), multiple system atrophy of the cerebellar type (yellow), and Parkinson's disease (red)

was assessed using an electrophoretic analysis (*Fig. 1B*, inset). The PAGE results confirmed the removal of the major impurity proteins after the hydrophobic chromatography step.

Final purification of α-synuclein to separate possible covalent and non-covalent dimers was performed using size exclusion chromatography. The chromatographic elution profile of α-synuclein (*Fig. 2*) indicated that the protein retention time on the column was 12.5 min, which corresponded to its expected monomeric weight. Additional peaks did not contain a polypeptide component and corresponded to conductivity variations induced by the α-synuclein buffer components. Electrophoretic analysis of fractions 1–3 revealed additional upper bands, whereas fraction 4 was of the highest purity and was used for RT-QuIC.

The content of aberrant α-synuclein in patients' CSF was measured using a modified RT-QuIC protocol [16]. During the measurement, pathological α-synuclein aggregates were partially denaturated by periodic quaking. With an excess of the recombinant monomeric protein in the reaction mixture, the misfolded aggregated protein, which binds to ThT, is amplified and enriched, which leads to an increase in the fluorescence intensity. The results are shown in *Figure 3*.

Sample 13 (stiff-person syndrome, SPS) had no significant increase in fluorescence, whereas samples 40 (multiple system atrophy, MSA) and 225 (Lewy body dementia, LBD) showed a fair increase in the fluorescence signal in the 40–60 h interval, which reached 145,000 and 120,000 RFU, respectively. Sample 7 (Parkinson's disease) showed the highest fluorescence signal of 170,000 RFU and the highest rate at reaching the plateau in an interval of 32–45 h. Thus, the analysis of RT-QuIC curves revealed an increase in the fluorescence level to 120,000–150,000 RFU in CSF samples from MSA and LBD patients, respectively, after about 75 h of observation and 170,000 RFU from PD after 55 h of observation. There was no increase in fluorescence in the CSF sample from the SRS patient, indicating the lack of aberrant α-synuclein in the biomaterial. A cerebrospinal fluid analogue (100 mM NaCl, 2 mM KCl, 1 mM $CaCl_2$, 5 mM urea, 300 μg/mL BSA, 2.5 mM glucose, $NaHCO_3$, pH 7.3) was used as a negative control.

DISCUSSION

The presented results indicate that the produced recombinant monomeric α-synuclein is the optimal substrate for use in the RT-QuIC technology. Previously, a variety of purification methods and extraction buffer compositions had shown to significantly affect the conformation, stability, and aggregation of α-synuclein, which, in turn, significantly complicates the interpretation of the assay results [17, 18]. In addition, standardization of methods for the production and purification of recombinant α-synuclein for diagnostic purposes is topical. By using the sequential combination of ion exchange and hydrophobic chromatography, we produced a highly purified target protein (⩾97% purity) which corresponds to recombinant monomeric α-synuclein and that may be successfully used in further functional tests. Such high purity and homogeneity of α-synuclein are the key factors in maintaining the reproducibility of laboratory tests for synucleinopathies. In the future, this will ensure greater reliability of RT-QuIC data and facilitate the jump from laboratory results to clinical practice.

Diagnosis of synucleinopathies is a complex task that requires high accuracy. In this pilot study, we used the RT-QuIC assay to analyze samples from patients with moderate LBD and MSA of the cerebellar type (MSA–M) who had impaired cognitive status. Samples derived from patients with stiff-person syndrome were used as a negative control. It should be noted that MSA can be diagnosed with 100% accuracy only postmortem, because its clinical picture overlaps with that of other synucleinopathies. There are two main subtypes, parkinsonian (MSA–P) and cerebellar (MSA–C), of this rare and rapidly progressing neurodegenerative disease. Clinical presentation of MSA–P includes symptoms typical of classical parkinsonism, while MSA–C is characterized by cerebellar ataxia. Similarly, the diagnosis of LBD is complicated by overlapping symptoms. LBD is often misdiagnosed as Alzheimer's disease due to the similarity of clinical manifestations. In this case, the only reliable diagnostic marker is the detection of aberrant α-synuclein in the patient.

The results of this study indicate that the onset of significant synuclein aggregation begins 50 h after the mixing of CSF samples with reaction mixture components. But in similar studies, the rise of aggregation curves started after approximately 12 h of incubation [19]. The rate of α-synuclein aggregation can depend on many factors, and the degree of purification is obviously the most critical one. Oligomers that could not be eliminated during recombinant protein purification can induce aggregation in the same way as natural pathological variants introduced into the reaction mixture do. On the one hand, a higher degree of α-synuclein purification increases the assay duration; on the other hand, it significantly reduces the background noise and allows for ex-

tended incubation times, which are essential for the reliable detection of synucleinopathies such as MSA, which are known for their weak seeding activity in the RT-QuIC assay [19].

CONCLUSION

Early detection of the aberrant proteins that are involved in neurotoxicity/neuroinflammation mechanisms and present in the systemic bloodstream and CSF is currently considered as one of the most critical frontiers in neuroscience research [20]. A series of pilot studies to quantify pathological aggregates in CSF from patients with synucleinopathies using RT-QuIC and highly purified recombinant α-synuclein have demonstrated the significant potential of this approach in advancing laboratory-based neurodiagnostics. Further research will need to be focused on the development of a standardized method for detecting pathological forms of α-synuclein that is informative, highly sensitive, specific, reproducible, and user-friendly, ensuring its suitability for future implementation in clinical practice. ●

This study was supported by a grant of the Ministry of Education and Science of the Russian Federation for conducting major research projects in priority areas of scientific and technological development, agreement No. 075-15-2024-638.

The authors are grateful to V.I. Muronets (Department of Animal Cell Biochemistry, Belozersky Institute of Physico-Chemical Biology) for providing the plasmid encoding α-synuclein.

REFERENCES

1. Gai W.P., Pountney D.L., Power J.H.T., Li Q.X., Culvenor J.G., McLean C.A., Jensen P.H., Blumbergs P.C. // Exp. Neurol. 2003. V. 181. № 1. P. 68–78. doi: 10.1016/s0014-4886(03)00004-9.
2. Spillantini M.G., Schmidt M.L., Lee V.M.-Y., Trojanowski J.Q., Jakes R., Goedert M. // Nature. 1997. V. 388. № 6645. P. 839–840. doi: 10.1038/42166.
3. Tong J., Wong H., Guttman M., Ang L.C., Forno L.S., Shimadzu M., Rajput A.H., Muenter M.D., Kish S.J., Hornykiewicz O., et al. // Brain. 2010. V. 133. № 1. P. 172–188. doi: 10.1093/brain/awp282.
4. Ma J., Gao J., Wang J., Xie A. // Front. Neurosci. 2019. V. 13. P. 552. doi: 10.3389/fnins.2019.00552.
5. Goedert M. // Science. 2015. V. 349. № 6248. P. 1255555. doi: 10.1126/science.1255555.
6. Call T., Ottolini D., Negro A., Brini M. // J. Biol. Chem. 2012. V. 287. № 22. P. 1791417929. doi: 10.1074/jbc.M111.302794.
7. Longhena F., Faustini G., Missale C., Pizzi M., Spano P., Bellucci A. // Neural Plast. 2017. V. 2017. P. 5012129. doi: 10.1155/2017/5012129.
8. Eliezer D., Kutluay E., Bussell R., Browne G. // J. Mol. Biol. 2001. V. 307. № 4. P. 10611073. doi: 10.1006/jmbi.2001.4538.
9. Winner B., Jappelli R., Maji S.K., Desplats P.A., Boyer L., Aigner S., Hetzer C., Loher T., Vilar M., Campioni S., et al. // Proc. Natl. Acad. Sci. USA. 2011. V. 108. № 10. P. 4194–4199. doi: 10.1073/pnas.1100976108.
10. Housmans J.A.J., Wu G., Schymkowitz J., Rousseau F. // FEBS J. 2023. V. 290. № 3. P. 554–583. doi: 10.1111/febs.16312.
11. Karpowicz R.J., Trojanowski J.Q., Lee V.M.-Y. // Lab. Invest. 2019. V. 99. № 7. P. 971981. doi: 10.1038/s41374-019-0195-z.
12. Rhoads D.D., Wrona A., Foutz A., Blevins J., Glisic K., Person M., Maddox R.A., Belay E.D., Schonberger L.B., Tatsuoka C., et al. // Neurology. 2020. V. 95. № 8. P. e1017–e1026. doi: 10.1212/WNL.0000000000010086.
13. Srivastava A., Alam P., Caughey B. // Biomolecules. 2022. V. 12. № 4. P. 576. doi: 10.3390/biom12040576.
14. Candelise N., Schmitz M., Thune K., Cramm M., Rabano A., Zafar S., Stoops E., Vanderstichele H., Villar-Pique A., Llorens F., et al. // Transl. Neurodegener. 2020. V. 9. № 1. P. 5. doi: 10.1186/s40035-019-0181-9.
15. Vascellari S., Orru C.D., Caughey B. // Front. Aging Neurosci. 2022. V. 14. P. 853050. doi: 10.3389/fnagi.2022.853050.
16. Okuzumi A., Hatano T., Fukuhara T., Ueno S., Nukina N., Imai Y., Hattori N. // Methods Mol. Biol. 2021. V. 2322. P. 3–16. doi: 10.1007/978-1-0716-1495-2_1.
17. Stephens A.D., Matak-Vinkovic D., Fernandez-Villegas

A., Kaminski Schierle G.S. // Biochemistry. 2020. V. 59. № 48. P. 4563–4572. doi: 10.1021/acs.biochem.0c00725. doi: 10.3390/biom12020324.
18. Al-Azzani M., Konig A., Outeiro T.F. // Biomolecules. 2022. V. 12. № 2. P. 324. doi: 10.3390/biom12020324.
19. Rossi M., Candelise N., Baiardi S., Capellari S., Giannini G., Orru C.D., Antelmi E., Mammana A., Hughson A.G., Calandra-Buonaura G., et al. // Acta Neuropathol. (Berl.). 2020. V. 140. № 1. P. 49–62. doi: 10.1007/s00401-020-02170-6.
20. Morris H.R., Lees A.J. // JAMA Neurol. 2024. V. 81. № 9. P. 905–906. doi: 10.1001/jamaneurol.2024.2381.

Two Key Substitutions in the Chromophore Environment of mKate2 Produce an Enhanced FusionRed-like Red Fluorescent Protein

D. A. Ruchkin[1], A. S. Gavrikov[1], D. V. Kolesov[1], A. Yu. Gorokhovatsky[1], T. V. Chepurnykh[1], A. S. Mishin[1], E. G. Maksimov[2], N. V. Pletneva[1], V. Z. Pletnev[1], A. M. Pavlova[1,3], V. A. Nikitin[1,2], A. M. Bogdanov[1,4*]

[1]Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Moscow, 117997 Russia
[2]Faculty of Biology, M.V. Lomonosov Moscow State University, Moscow,119992 Russia
[3]Pirogov Russian National Research Medical University, Moscow, 117997 Russia
[4]Department of Photonics, İzmir Institute of Technology, İzmir, 35430 Turkey
*E-mail: noobissat@ya.ru

Received: October 20, 2024; in final form, February 19, 2025
DOI: 10.32607/actanaturae.27545

ABSTRACT **Red fluorescent proteins (RFPs) are often probes of choice for living tissue microscopy and whole-body imaging. When choosing a specific RFP variant, the priority may be given to the fluorescence brightness, maturation rate, monomericity, excitation/emission wavelengths, and low toxicity, which are rarely combined in an optimal way in a single protein. If additional requirements such as prolonged fluorescence lifetime and/or blinking ability are applied, the available repertoire of probes could dramatically narrow. Since the entire diversity of conventional single-component RFPs belongs to just a few phylogenetic lines (DsRed-, eqFP578- and eqFP611-derived being the major ones), it is not unexpected that their advantageous properties are split between close homologs. In such cases, a systematic mutagenetic analysis focusing on variant-specific amino acid residues can shed light on the origins of the distinctness between related RFPs and may aid in consolidating their strengths in new RFP variants. For instance, the protein FusionRed, despite being efficient in fluorescence labeling thanks to its good monomericity and low cytotoxicity, has undergone considerable loss in fluorescence brightness/lifetime compared to the parental mKate2. In this contribution, we describe a fast-maturing monomeric RFP designed semi-rationally based on the mKate2 and FusionRed templates that outperforms both its parents in terms of molecular brightness, has extended fluorescence lifetime, and displays a spontaneous blinking pattern that is promising for nanoscopy use.**

KEYWORDS **RFP, FusionRed, mKate2, fluorescent protein, fluorescence lifetime.**

ABBREVIATIONS **FP – fluorescent protein; RFP – red fluorescent protein; FLIM – fluorescence lifetime imaging microscopy; BALM – bleaching/blinking assisted localization microscopy; OSER – organized smooth endoplasmic reticulum; NE – nuclear envelope; PEI – polyethyleneimine; SMLM – single molecule localization microscopy.**

INTRODUCTION

Current bioimaging techniques recruit a vast diversity of fluorescent probes; among those, genetically encoded fluorophores such as fluorescent proteins (FPs) are in favor. FPs enable highly specific intracellular labeling, live-cell super-resolution, fluorescence lifetime imaging microscopy (FLIM), etc. [1, 2]. In turn, red fluorescent proteins (RFPs), a polyphyletic group [3–6] of anthozoan FPs emitting in the red region of the spectrum, are of particular relevance in whole-body and/or deep-tissue imaging, owing to their enhanced detectability within the optical transparency window characterizing the local absorption minimum of animal tissues at wavelengths of ~ 600–1200 nm [7–9].

Among existing RFP variants, FusionRed [10] is often the probe of choice for live-cell imaging (including the visualization of fine subcellular structures) thanks to its "supermonomericity," i.e., an ability to maintain a highly monomeric state even at the

high local concentrations that are typical of specialized localizations within mammalian cells [11, 12], as well as its low acid sensitivity and toxicity. It is thus supposed to be used as a probe fused to the proteins of interest without affecting their natural activities and spatial structures, or can function as a fluorescent core of genetically encoded indicators [13–16]. Along with this, there are definite drawbacks that take from the practical value of FusionRed as a multipurpose fluorescence tag and call for further improvement of this RFP. In that perspective, the issue of the modest molecular brightness of FusionRed has been addressed in the elegant work by Jimenez lab, where both directed evolution [17] and semi-rational design [18] were utilized to engineer brighter variants of FusionRed (specifically, FusionRed-MQV [19] has an ~ fourfold higher molecular brightness over the parental RFP, although its emission peak comes with a 20-nm hypsochromic shift). The high-resolution spatial structure of FusionRed revealed that almost half of its molecules carry an immature chromophore [20]; this feature impinges on the effective brightness (well under the level expected based on the measured molecular brightness) of FusionRed as a fluorescence probe, suggesting further room for improvement via a structure-based design of daughter RFP variants.

Importantly, FusionRed is a descendant of the mKate2 protein [21] that emits at 633 nm (its emission maximum is 25-nm red-shifted relative to that of FusionRed) and is currently the brightest monomeric far-red FP. FusionRed differs from mKate2 in 17 amino acid substitutions introduced semi-rationally through several consecutive rounds of mutagenesis [10]: Hence, there is no consolidated picture that describes the particular role of every substitution, specifically since the structural foundations of the spectral differences (including the extinction coefficient, fluorescence quantum yield and lifetime, as well as the excitation/emission maxima positions) between the FusionRed and mKate2 proteins are not clear enough. Based on an analysis of the spatial structure of FusionRed [20], one can assign an essential role to three residues in the chromophore environment: Arg/Lys-67, Cys/Ala-158, and His/Arg-197 (FusionRed/mKate2, respectively). Here, we systematically studied the influence of these residues on the properties of both proteins through exhaustive reciprocal site-directed mutagenesis. Among the representatives of the library obtained, there is one that is remarkable: mKate2-K67R/R197H, which shows a striking similarity in its steady-state absorption and fluorescence spectra to those of FusionRed and is 2.2-fold brighter than the latter. This RFP inherits the advantages of both sister proteins; namely, it demonstrates a monophasic fluorescence decay similarly to mKate2 and performs well as a fusion tag like FusionRed. Interestingly, purified mKate2-K67R/R197H possesses a well-marked pattern of spontaneous fluorescence blinking that might be promising for use in super-resolution microscopy.

MATERIALS AND METHODS

Site-directed mutagenesis

A modified IVA-cloning [22] procedure was applied to produce the site-specific mutants of mKate2 and FusionRed. The genes of the chosen RFPs, cloned into the pQE-30 vector backbone (Qiagen, Germantown, Maryland, USA) using BamHI/HindIII endonuclease sites, were utilized as primary templates. The forward oligonucleotides were designed to have a 5'-terminal 15– to 20-nt length region homologous to the template DNA (needed to provide bacterial recombination), followed by a triplet with the mutation of interest and a 3'-terminal priming region designed to anneal at 60°C. The reverse oligonucleotides consisted of a similar recombination-guide part of 15–20 nt and a 3'-terminal priming sequence; both made up an annealing temperature of 60°C when possible. In cases of higher calculated annealing temperatures, the 5'-end fragment was considered partially annealing. The 3' and 5' terminal bases of both primers were selected in such a way as not to pair complementarily in order to avoid self-annealing of long oligonucleotides when possible; simultaneously, the terminal 3'-nucleotides of both primers were selected, by design, to form a strong complementary pair with the template sequence. The reverse primer could never anneal to the forward one with 3'-terminus resulting in a blunt-end. PCR was performed using the standard Phusion Polymerase (ThermoFisher, Waltham, Massachusetts, USA) protocol and lasted 35 cycles; the template DNA made up for a total of 50 ng per reaction. The primers used had the following sequences:

(a) FusionRed-R67K:

Forward – 5'-agcttcatgtacggcagcaaaaccttcatcaagcaccctccgg-3'

Reverse – 5'-gctgccgtacatgaagctggtag-3'

(b) FusionRed-C158A:

The mutant was engineered in the previous study [20].

Forward – 5'-cggcggcctggaaggcgcagcagacatggccctgaagctcg-3'

Reverse – 5'-tgcgccttccaggccgccgtcagcggggtacatcgtctcg-3'

(c) FusionRed-H197R:

Forward – 5'-ggcgtctacaacgtggacagaagactggaaagaatcaaggaggc-3'

Reverse – 5'-gtccacgttgtagacgccgggcatcttgaggttcgtagcg-3'
(d) mKate2-K67R:
Forward – 5'-agcttcatgtacggcagcagaaccttcatcaaccacacccaggg-3'
Reverse – 5'-tgctgccgtacatgaagctggtag-3'
(e) mKate2-A158C:
The mutant was engineered in the previous study [21].
Forward – 5'-ggcctggaaggcagatgcgacatggccctgaagctcg-3'
Reverse – 5'-tctgccttccaggccgccgtcagcggggtacag-3'
(f) mKate2-R197H:
Forward – 5'-ggcgtctactatgtggaccacagactggaaagaatcaaggaggc-3'
Reverse – 5'-gtccacatagtagacgccgggcatcttgaggttcttagcg-3'

The PCR products were reprecipitated and treated with DpnI restriction endonuclease to remove the initial template DNA. For transformation (needed for construct assembly), 700 ng of the PCR product was taken per 100 μL of the aliquot of *E. coli* XL1-Blue competent cells (Evrogen, Moscow, Russian Federation).

Protein expression and purification

The FP variants were expressed in the *E. coli* XL1-Blue strain for 72 h at 37°C. After centrifugation, bacterial biomass was resuspended in PBS (GIBCO, ThermoFisher Scientific, Waltham, Massachusetts, USA), pH 7.4, and treated with ultrasound using a Sonics Dismembrator (Fisher Scientific, Pittsburgh, Pennsylvania, USA). The proteins were then purified using the TALON metal-affinity resin (Clontech, Mountain View, California, USA) added earlier and washed in PBS according to the manufacturer's protocol, solubilized using 0.3 mM imidazole (pH 8.0). The protein eluates were then desalted and concentrated by ultrafiltration with Amicon Ultra 0.5 10K columns (Merck Millipore, Burlington, Massachusetts, USA). The resulting concentrated protein solution (typically ~5 mg/mL) was ready for use in a SDS-PAGE analysis or spectroscopy or could be stored for a short time at 4°C until use.

Steady-state absorption and fluorescence spectroscopy

The absorbance and fluorescence spectra were recorded using a Cary100 UV/VIS spectrophotometer and a Cary Eclipse fluorescence spectrophotometer (Agilent Technologies, Santa Clara, California, USA), respectively. A protein solution in PBS (pH 7.4) was used in all the cases. The fluorescence quantum yields and extinction coefficients were determined as described earlier [20].

Monomericity testing

Gel filtration. Gel filtration experiments were performed using a Superdex® 200 Increase 10/300 GL column (Cytiva, Uppsala, Sweden) equilibrated with a 20 mM sodium phosphate buffer (pH 7.4) containing 150 mM NaCl at 24°C at a flow rate of 0.75 mL/min. The column was connected to an Agilent 1260 Bio-Inert LC system equipped with an in-line Agilent 1260 diode array detector and an Agilent 1260 fluorescence detector and calibrated using cytochrome C (12.4 kDa), carbonic anhydrase (29 kDa), bovine serum albumin (66 kDa), alcohol dehydrogenase (150 kDa), b-amilase (200 kDa), and ferritin (450 kDa). The calibration details are presented in *Fig. S6* and *Table S1*. The equipment was controlled by the Agilent OpenLAB CDS ChemStation Edition C.01.07 SR3 software.

OSER assay. The OSER assay was undertaken in two variants. The first one, based on HeLa cells, was similar to that described in ref. [12]. The cells were transfected with the FuGENE® HD Transfection Reagent (Promega, Woods Hollow Road, Madison, USA) according to the commercial protocol. Images were acquired with wide-field fluorescence microscopy using a modified Leica 6000LX inverted microscope equipped with an mCherry filter cube (see the Widefield fluorescence microscopy section). The images were processed using Fiji ImageJ distribution (version 2.9.0/1.54b). Whorl-like structures were then identified according to the guidelines elaborated by Constantini et al. [23]. Due to the lack of whorl-like structures in more than 80% of the transfected HeLa cells and the resulting considerable difficulty in capturing enough whorl-possessing HeLa cells for a valid statistical analysis, the ratios between the mean fluorescent intensities of the nuclear envelope and whorl-like structures were not calculated.

The second variant of the OSER assay was similar to that described in ref. [23]. U2OS cells were transfected with polyethyleneamine (PEI, Sigma-Aldrich, Saint Louis, Missouri, USA). The observation was performed 18 h post-transfection; the images were analyzed using the same Fiji ImageJ software to extract the mean NE and OSER signals. Not less than three linear ROIs for NE were traced for each cell using the "straight" tracing instrument; the "freehand" instrument was used for OSER ROIs. The ratios were calculated using the GraphPad Prism10 software.

Engineering of mammalian constructs. Mammalian expression plasmids encoding fusions of Diogenes with vimentin (vimentin-Diogenes), lifeact (life-

act-Diogenes), ensconsin (ensconsin-Diogenes) and cytokeratin (Diogenes-cytokeratin), as well as the fusion with the cytoplasmic end of an endoplasmic reticulum signal anchor membrane protein (CytERM; used in the OSER assay), were assembled using Golden Gate cloning according to the MoClo standard procedure [24–26]. Each transcriptional unit for mammalian expression included a CMV promoter, a coding sequence for the fusion protein, and the SV40 terminator. All Golden Gate cloning reactions were performed in the T4 ligase buffer (SibEnzyme, Moscow, Russia) with 10 U of T4 ligase, 20 U of either the BsaI or BpiI restriction endonuclease (ThermoFisher, Waltham, Massachusetts, USA), and 100–200 ng of the DNA of each fragment. The assembly reactions were performed under the following conditions: 30 cycles of 37°C and 16°C incubation (90 s at 37°C, 180 s at 16°C).

Widefield fluorescence microscopy

Widefield fluorescence microscopy was performed with a Leica 6000LX inverted microscope equipped with a Leica HCX PL APO 100X/1.40–0.70NA oil immersion objective, a Zyla sCMOS camera (Andor, Oxford, UK), and a CoolLED pE-300 light source. An mCherry cube filter set (Leica, Wetzlar, Germany) was used (excitation filter: 560/40, emission filter: 630/75). The typical illumination power ranged from 1 to 5 W/cm^2 with exposure times ranging from 50 to 150 ms.

pH-stability measurement

A set of pre-made buffer solutions with pH ranging from 3 to 10.55 was used to prepare the protein samples; the solutions contained 130 mM KCl, 30 mM NaCl, 0.5 mM $MgCl_2$, 0.2 mM EGTA, and 30 mM HCl–$NaH_2C_6H_5O_7$ (pH 3.0–4.5) or 15 mM KH_2PO_4–Na_2HPO_4 (pH 5.0–7.5) or 20 mM $Na_2B_4O_7$–HCl/NaOH (pH 8.0–11.0) [27]. Each probe contained 5 μg/mL of the purified and desalted RFP. For each sample, the emission spectra were measured with a Cary Eclipse Fluorescence Spectrometer twice for each of three temporary points (immediately after preparation, after 3 min and 5 min), with a total of six measurements per sample in a spectral range from 560 nm to 700 nm at λ_{ex} = 540 nm using a 5 nm ex/em slit, an equal photomultiplier (PMT) voltage, and the scanning speed values. The fluorescence intensity values at the emission maxima were averaged from six reads. The averaged data from all pH points for each RFP were normalized to the maximum value within the set and plotted on a graph with standard deviations. The sigmoidal regions of the graphs were fitted (4PL logistic curve, 95% confidence, n = 6) in GraphPad Prism10; the p*K*a of each protein was defined at a read of 0.5 on the fitting curve.

Fluorescence lifetime measurements

Nanosecond and picosecond setups. Measurements were made using a time-resolved (TCSPC) miniTau fluorescence spectrometer (Edinburgh Instruments, Livingston, UK) in a 20 ns window divided into 2,048 time channels. The fluorescence was excited using: (i) an EPL-450 picosecond laser (Edinburgh Instruments, Livingston, UK) with a central emission wavelength of 445.6 nm, a pulse width (FWHM) of ~ 90 ps@10 MHz driven at a repetition rate of 20 MHz; (ii) an EPLED-590 nanosecond pulsed LED (Edinburgh Instruments, Livingston, UK) with a central emission wavelength of 590 nm and a pulse width (FWHM) of ~ 1.3 ns driven at a repetition rate of 20 MHz. The photons were counted in a spectral range of 575–625 nm. The data processing, visualization, and determination of χ^2 (Pearson's test) were performed using the Fluoracle 2.5.1 software (Edinburgh Instruments, Livingston, UK).

Femtosecond setup. The fluorescence decay kinetics of RFPs were recorded by a single-photon counting (SPC) detector with an ultra-low dark count rate (HPM-100-07C, Becker & Hickl, Germany) in the 620/10 spectral window and adjusted by an ML-44 monochromator (Solar, Belarus). Fluorescence was excited at 590 nm (repetition rate, 80 MHz; pulse width, 150 fs; optical power, 5 mW) using the second harmonics (ASG-O, Avesta Project LTD, Moscow, Russia) of a femtosecond optical parametric oscillator (TOPOL-1050-C, Avesta Project LTD.) pumped by a Yb femtosecond laser (TEMA-150, Avesta Project LTD). The emission signal was collected perpendicular to the excitation beam. The sample temperature was stabilized during the experiment at 25°C with a cuvette holder (Qpod 2e) with a magnetic stirrer (Quantum Northwest, USA). The SPCM Data Acquisition Software v. 9.89 (Becker & Hickl, Germany) was used for data acquisition. The SPCImage software (Becker & Hickl, Germany) was used for the exponential fitting of fluorescence decays considering the incomplete decay of RFPs due to the high repetition rate. Post-processing and visualization of the collected data were performed using the Origin Pro 2018 software (OriginLab Corporation, USA).

Photostability measurements

Purified proteins, low excitation intensity. For the photobleaching experiments, the RFPs immobilized on TALON metal-affinity resin beads were imaged. Measurements were performed using a DMIRE2 TCS SP2 laser scanning confocal inverted microscope

(Leica Microsystems, Wetzlar, Germany) equipped with an HCX PL APO lbd.BL 63× 1.4NA oil objective and a 1.2 mW HeNe laser. The red fluorescent signal was acquired using the 543 nm excitation laser line and detected within a 560–670 nm spectral range. The selected field of view (16× zoom) was scanned in the time-lapse (between frames) mode, wherein the sequence of detection and bleaching frames was repeated 500–1,500 times without delay. To detect the red fluorescence signal, a 10%–20% laser power and PMT voltage of 700–800 V were used. For fluorophore photobleaching, the 100 % laser power (yielding about 2 W/cm^2 power density) was used. The fluorescence data were all background-subtracted, averaged (n = 5), and normalized to the maximum value. A LaserCheck (Coherent, Saxonburg, Pennsylvania, USA) power meter was used to measure the total power of the excitation light after the microscope's objective. The light power density (W/cm^2) was estimated by dividing the total power by the area of the laser-scanned region.

In cellulo measurements, high excitation intensity. For the photobleaching experiments, the fluorescence signal of the RFPs, transiently expressed in the HeLa cell culture and lacking a specific intracellular targeting signal, was acquired. Measurements were performed using a Nanoimager S (ONI, Oxford, UK) microscope equipped with an Olympus UPlanSApo ×100 NA 1.40 oil immersion objective, a 561 nm laser, a 560 nm on-camera beam splitter, and a Scope8 sCMOS camera. The cells were irradiated in the epi-fluorescence mode with the 561 nm laser at a power density of 800 W/cm^2 with simultaneous continuous signal recording and minimal delays between frames. Data analysis was performed using the FiJi ImageJ 1.53f51 software [28].

Single-molecule localization microscopy

Super-resolution BALM imaging of the cytoskeleton of cultured mammalian cells was performed as follows. Immediately before imaging, the cell medium was replaced with the minimal essential medium (MEM, Sigma-Aldrich, Saint Louis, Missouri, USA) supplemented with 20 mM HEPES. Single-molecule localization super-resolution imaging of living cells was performed using a Nanoimager S (ONI, Oxford, UK) microscope equipped with an Olympus UPlanSApo ×100 NA 1.40 oil immersion objective, a 561 nm laser, a 560 nm on-camera beam splitter, and a Scope8 sCMOS camera. Imaging was performed using the following imaging condition set: a 2 kW/cm^2 561 nm laser and 16.7 ms frame time (60 fps acquisition speed). The imaging procedure with additional photostimulation by 405-nm laser flashes had the following conditions: imaging by a 561 nm laser operating at 2 kW/cm^2 was accompanied by 405 nm laser flashes with a duration of 0.4 s and an illumination density of ~ 215 W/cm^2, applied every 22 s. The frame recording time was 16.7 ms, and the acquisition speed was 60 fps. The difference between the signal-to-noise ratio of mKate2-K67R/R197H, TagRFP-T, and mKate2 localizations was assessed using the Kolmogorov–Smirnov test. Image acquisition and super-resolution reconstruction were performed using the NimOS 490 1.18.3.15066 software (ONI, Oxford, UK). Image reconstruction was done using default parameters. Data analysis was performed using the FiJi ImageJ 1.53f51 [29] and custom Python 3.9 scripts.

RESULTS AND DISCUSSION

To clarify the roles played by particular amino acid substituents in the chromophore environment of FusionRed and mKate2 in the physicochemical distinctness of these fluorescent proteins (including their spectral differences and features of chromophore maturation), we conducted a systematic mutational analysis implying the introduction of single, double, and triple reciprocal substitutions (*Fig. S1*) at the key positions 67, 158, and 197, which had earlier been identified as 'gatekeepers' of the FusionRed chromophore behavior based on its crystal structure [20].

Description of the reciprocal mutants

Single mutations. Substitution at position 67 (Arg↔Lys) had differential effects on the parental proteins. Thus, the mKate2-K67R variant was found to have negligible absorption in the visible range and to be almost non-fluorescent; the mutation probably strongly affected folding and/or chromophore maturation. Conversely, FusionRed-R67K contains several well-marked spectral species, which likely correspond to different chromophore structures (*Table 1*, *Fig. S2*). Its absorption (being simultaneously fluorescence excitation) peaked at 389, 514, and 580 nm. The latter red emissive species ($\lambda_{abs/ex}$ = 580 nm, λ_{em} = 610 nm) behaves similarly to the parental FusionRed, while both short-wave species are supposed to be the populations of immature chromophore.

We assumed that a blue-emitting spectral species ($\lambda_{abs/ex}$ = 389 nm, λ_{em} = 450 nm) corresponds to the neutral GFP-type chromophore, which is the well-described intermediate of the DsRed chromophore maturation [30, 31]. A yellow fluorescent species ($\lambda_{abs/ex}$ = 514 nm, λ_{em} = 522 nm) of FusionRed-R67K, which spectrally resembles conventional yellow fluorescent proteins (EYFP, TagYFP) that bear a GFP-

Table 1. Summary of the spectral properties, chromophore maturation, and post-translational chemistry observed in the set of single, double, and triple reciprocal mutants of FusionRed and mKate2

Protein	Absorption peak, nm	$\lambda_{ex}/\lambda_{em}$, nm	EC[a] ($M^{-1}\cdot cm^{-1}$)	FQY[b]	Molecular brightness (EC · QY/1000)	Comment
FusionRed-R67K/C158A/H197R	376; 488; 583	376/409; 488/508; 583/616	n/d	<0.05	n/d	poor maturation*
FusionRed-C158A/H197R	n/d	n/d	n/d	n/d	n/d	poor maturation*
FusionRed-R67K/H197R	**380; 488; 584**	**380/449; 488/511; 584/616**	**n/d**	**0.62**	**n/d**	
FusionRed-R67K/C158A	386; 513; 580	n/d	n/d	n/d	n/d	poor maturation*
FusionRed-H197R	n/d	570/607	n/d	<0.01	n/d	poor maturation*
FusionRed-C158A	**571**	**571/598**	**91 000**	**0.24**	**21.84**	**[20]**
FusionRed-R67K	**389; 514; 580**	**389/450; 514/522; 580/610**	**n/d**	**0.3**	**n/d**	
FusionRed	**580**	**580/608**	**94 500**	**0.19**	**17.9**	**[10]**
mKate2	**586**	**588/633**	**62 500**	**0.4**	**25**	**[21, 33]**
mKate2-K67R	405; 588	n/d	n/d	n/d	n/d	poor maturation*
mKate2-A158C	**380; 590**	**590/624**	**47 300**	**0.47**	**22.2**	**[20]**
mKate2-R197H	**385; 510; 582**	**510/520; 582/612**	**n/d**	**0.26**	**n/d**	
mKate2-K67R/A158C	n/d	n/d	n/d	n/d	n/d	poor maturation*
mKate2-K67R/R197H	**579**	**579/603**	**90 000**	**0.44**	**39.6**	
mKate2-A158C/R197H	**380; 513; 583**	**380/435; 583/611**	**n/d**	**0.39**	**n/d**	
mKate2-K67R/A158C/R197H	n/d	n/d	n/d	n/d	n/d	poor maturation*

Note: n/d – not determined.
[a]Molar extinction coefficient; EC has not been determined for the variants possessing several spectral species.
[b]Fluorescence quantum yield; FQY was measured for the red emissive species only.
*Label applied if a low-to-invisible bacterial biomass fluorescence 48 h post transformation and/or low relative absorbance at the chromophore-related wavelengths (e.g., A280/A580 > 10) was observed.

chromophore π-stacked with the tyrosine-203 residue [32], is less usual for RFPs. As one can speculate, the R67K substitution led to partial "freezing" of the FusionRed chromophore maturation at the pre-last oxidation step (GFP-like chromophore), and an anionic GFP-chromophore (usually absorbing at 470–500 nm) underwent a bathochromic spectral shift (to the yellow species) due to its π-stacking with the imidazole ring of histidine-197.

The influence of a reciprocal mutation at position 158 (Cys↔Ala) on the spectral properties of FusionRed and mKate2 has earlier been documented [20]. Although this substitution did not result in the formation of new spectral species or severe inhibition of the chromophore maturation (or strong changes in molecular brightness), we include the data on the corresponding mutants (FusionRed-C158A and mKate2-A158C) in *Table 1* for uniformity.

The mutations at position 197 (His↔Arg) had effects that were antagonistic to those of R67K/K67R. Similarly to mKate2-K67R, FusionRed-H197R was found to possess negligible absorption in the visible spectral region and to be almost non-fluorescent due to hindrance in the chromophore maturation. mKate2-R197H has its the absorption maxima at 385, 510, and 582 nm, of which two latter are also the fluorescence excitation peaks, and emits at 520 nm and 612 nm (*Table 1*, *Fig. S3*). The identities and origins of these spectral species are suggested to be the same as those of FusionRed-R67K. Notably, there is a well-marked hypsochromic shift in the absorption/emission maxima of the red species of mKate2-R197H

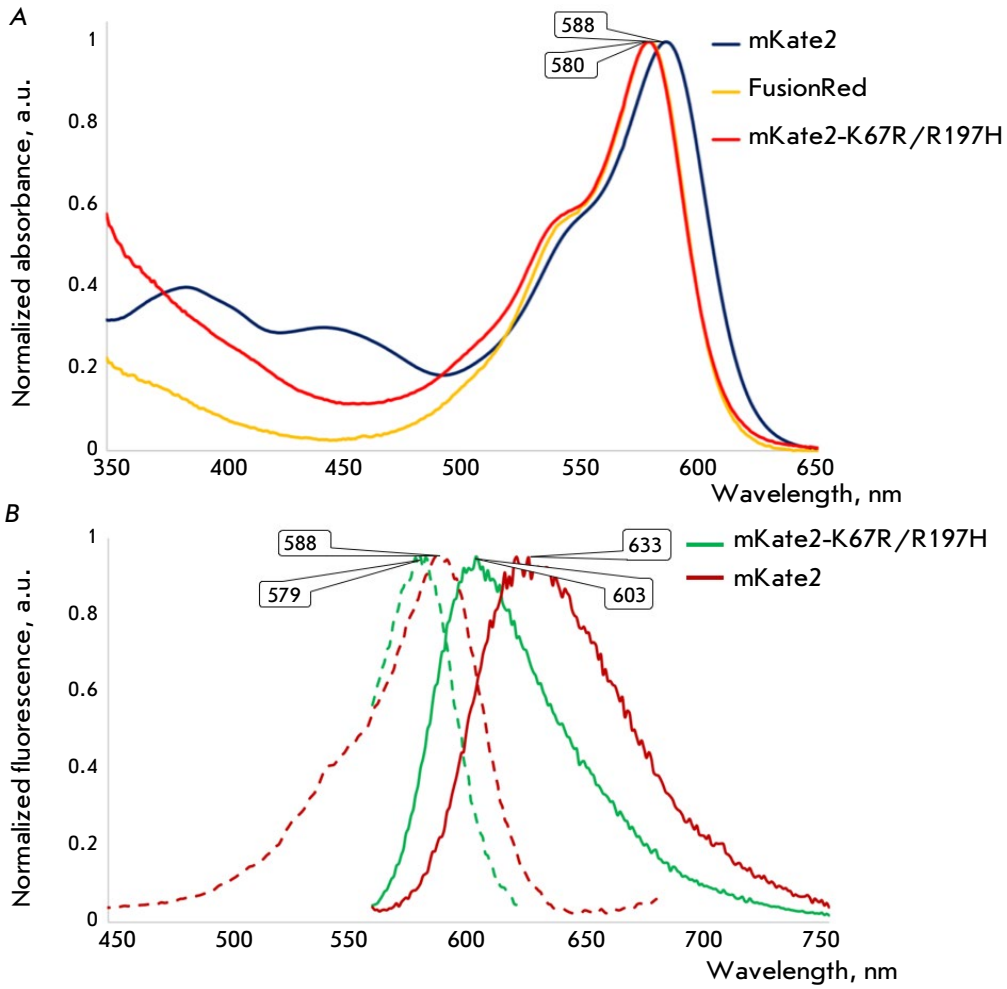


Fig. 1. The absorption (*A*) and fluorescence (*B*) spectra of mKate2-K67R/R197H compared to those of its parent mKate2 and close homologue, FusionRed (absorption only). The wavelengths of the maxima of major bands are shown in the bubbles. In the fluorescence graph, dashed lines show fluorescence excitation; solid lines show fluorescence emission

Table 2. Brief summary of the spectral properties possessed by mKate2, FusionRed, and mKate2-K67R/R197H, aka Diogenes

Protein	λ_{ex}, nm	λ_{em}, nm	EC ($M^{-1} \cdot cm^{-1}$)	FQY	Molecular Brightness (EC · FQY/1000)	Fluorescence lifetime, ns#
mKate2	588	633	62,500	0.4	25	2.4[1]/2.05[1]
FusionRed	580	608	94,500	0.19	17.955	1.6[2]
mKate2-K67R/R197H	579	603	90,000	0.44	39.6	2.2[1]

#Intensity-weighted average lifetimes are shown. For mKate2, two lifetime values obtained at different setups are shown (see the "Fluorescence lifetime" section).
[1]Monoexponential fitting gave adequate goodness ($\chi^2 \leq 1.3$);
[2]Biexponential fitting gave adequate goodness ($\chi^2 \leq 1.3$).

compared to the parental mKate2 protein (582/612 nm vs. 588/633 nm), which proves the key role played by His-197 in the determination of the FusionRed spectral distinction.

Double mutations. The chromophore maturation in both proteins was generally less tolerant to the introduction of sets of two amino acid substitutions. Thus, three out of six double mutants were either extremely dim and weakly absorbing or almost non-fluorescent and having no detectable absorption maxima in the visible range (see *Table 1*). In the relatively bright FusionRed-R67K/H197R variant, an antagonistic functionality of the residues at positions 67 and 197 is expressed. The R67K mutation partially unlocks the chromophore maturation strongly inhibited

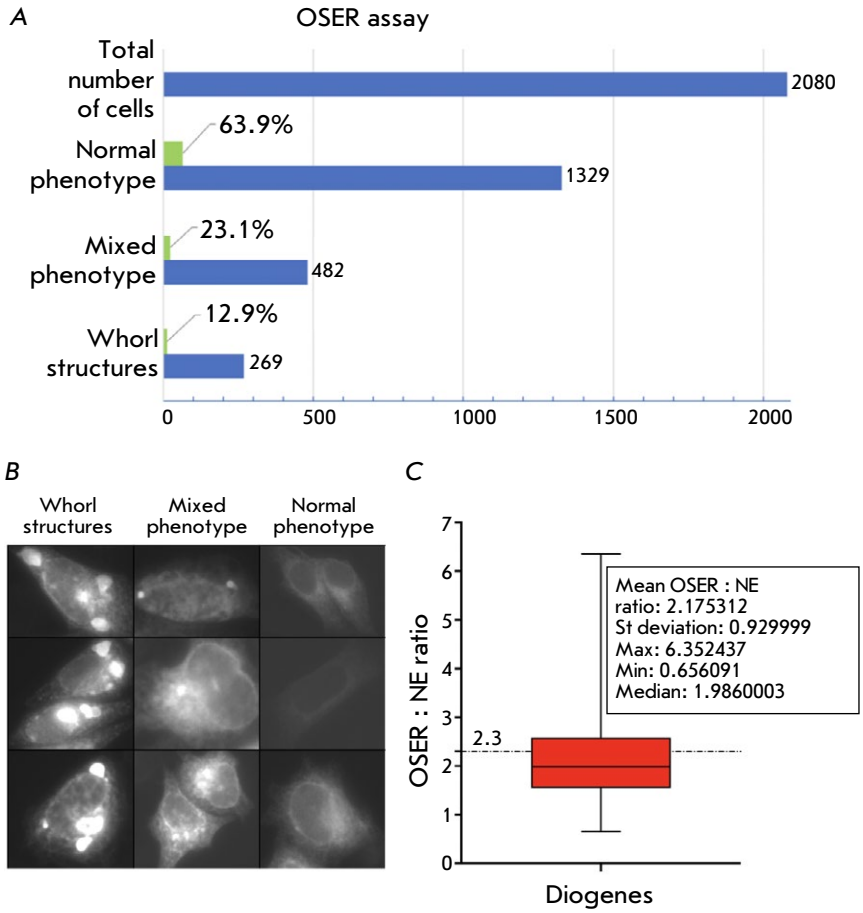


Fig. 2. Summary of the Diogenes examination using the OSER assay. (*A*) Histogram showing the total cell number (HeLa) and distribution of three phenotypes between them; (*B*) The gallery of fluorescence images illustrating the phenotypes observed during the OSER-based examination in live HeLa cells. The "Whorl structures" column depicts the cells where homo-oligomerization of the label yielded the typical organized smooth endoplasmic reticulum (OSER) structures (whorls) as they have been earlier described [12, 23]. The "Mixed phenotype" row represents the cases when some labeling artifacts other than typical whorls (small puncta, dots, regions with locally increased brightness) are observed. The "Normal phenotype" row exemplifies the cells with an evenly stained ER tubular network; (*C*) Graph showing the quantitative analysis of the OSER : NE intensity ratio in live U2OS cells. The red box indicates 25–75 percentile of the readings; whiskers stand for the min and max read of the set. Median is illustrated by a horizontal line within the box; the dash-and-dot line at 2.3 indicates the monomer's threshold OSER : NE value according to the original paper [23]. Descriptive statistics for the data are shown in the inset

by H197R. FusionRed-R67K/H197R possessed three emissive species (*Table 1*, *Fig. S4*): the blue-emitting neutral GFP ($\lambda_{abs/ex}$ = 380 nm, λ_{em} = 449 nm), the green-emitting anionic GFP ($\lambda_{abs/ex}$ = 488 nm, λ_{em} = 511 nm), and the red-emitting DsRed-like one ($\lambda_{abs/ex}$ = 584 nm, λ_{em} = 616 nm). Remarkably, the red form of FusionRed-R67K/H197R showed a well-defined bathochromic shift in both absorption and emission (4 and 8 nm, respectively) compared to the parental FusionRed, thus providing additional evidence of the role of the substituent at position 197 in spectral tuning of RFPs. The mKate2-A158C/R197H variant demonstrates a complex spectral behavior (*Table 1*, *Fig. S5*) similar to that observed for the FusionRed R67K and mKate2 R197H proteins. In mKate2-A158C/R197H, the R197H substitution is likely to provide a hypsochromic shift of the fluorescence spectra of the red species, as well as a stacking interaction with the immature "green" chromophore, leading to the formation of spectral species with an absorption maximum at 513 nm. Importantly, the latter was found to be non-fluorescent.

mKate2-K67R/R197H was the only variant from the library of the mKate2/FusionRed reciprocal mutants that exhibited fast chromophore maturation and high molecular brightness. Its fluorescence quantum yield of 0.44 and extinction coefficient of 90,000 make it 1.6 times brighter than mKate2 and 2.2 times brighter than FusionRed protein. In contrast to the parental protein, this double mutant lacks the minor shortwave absorption peaks at ~ 390 and ~ 450 nm attributed to immature chromophore and exhibits a blue-shifted main absorption band with a pronounced 'shoulder' at ~ 540 nm, typical of FusionRed (*Tables 1* and *2*, *Fig. 1*).

Triple mutations. The introduction of the full triad of reciprocal 67/158/197 substitutions had a striking effect on protein maturation. Thus, both triple mutants, mKate2-K67R/A158C/R197H and FusionRed-C158A/H197R/R67K, displayed undetectable-to-very-low absorbance/fluorescence in the visible part of the spectrum (see *Table 1*), presumably indicating chromophore "freezing" in the early stages of maturation. The information value of these variants in terms of establishing the molecular determinants of the spectral distinctiveness of FusionRed/mKate2 turned out to be low.

Over all, our phenotypic analysis revealed that the substitutions at position 67 likely cause a shift of the chromophore within the beta-barrel, since, after the mutagenesis, the spectroscopic signs of its π-stacking interaction with histidine-197 (if any) are changed compared to those in the parental proteins. The impact of these substitutions on chromophore maturation is also evident: in all the cases, except for mKate2-K67R/R197H, they led to either a strong alteration in the maturation or at least an elevated presence of the shortwave spectral species representing maturation intermediates. Substitutions at position 197 induce spectral shifts, a bathochromic absorption/emission shift in the case of the FusionRed-derived variants and a hypsochromic one in the mKate2 mutants. We speculate that the reason behind this phenomenon might have to do with the π-stacking interaction between the chromophore and histidine (occupying position 197 in the original FusionRed) switched off/on by reciprocal mutations. Additionally, these mutations were shown to have a noticeable impact on the chromophore maturation process.

Physicochemical properties of mKate2-K67R/R197H and its performance in microscopy

Next, we endeavored to achieve a detailed characterization of the physicochemical properties of the mKate2-K67R/R197H protein, named Diogenes, by focusing on its performance in cellular fluorescence imaging.

Oligomeric state and protein labeling. The first step was to analyze the protein's oligomeric state, which is among the key predictors of efficient low-disturbed/minimally invasive labeling of intracellular targets. The gel filtration chromatography data (*Figs. S6* and *S7*) show that the purified protein elutes as a single peak with an estimated molecular weight of ~38 kDa (at a concentration of up to at least 5 mg/mL). Since this molecular weight corresponds neither to the monomer (~25 kDa) nor to the dimer (~50 kDa), the gel filtration data cannot be interpreted unambiguously. One can assume that concentrated Diogenes in an aqueous solution is either a strict monomer or a strict dimer, having anomalous chromatographic mobility in either case. Alternatively, it is possible that we observed an equilibrium mixture of the monomeric and dimeric states.

It would be reasonable to expect that, in terms of their oligomeric state, Diogenes would be close to parental mKate2, which was originally described as a monomer [21], with further evidence of some propensity to oligomerize in aqueous solutions at a high concentration [10] and in cellulo [12]. Meanwhile, it is important to evaluate how the monomeric quality of this variant compares with that of its spectral analog, FusionRed. During the engineering of the latter, considerable effort was devoted to optimizing the outer surface of the beta-barrel, including the elimination of potentially dimerizing residues [10]. It was indeed shown that purified FusionRed behaves as a strict monomer [10] and scores higher on the monomericity ranking than mKate2 when examined in cellulo (91.5 ± 3.0% vs. 81.1 ± 6.1% in the OSER assay [12]). However, establishing a causal link between the monomerizing mutations introduced into FusionRed and its better performance in cellulo remains somewhat debatable, since the rational design of these substitutions was based on the spatial structure of mKate rather than on that of mKate2 [34]. Moreover, protein folding and the observed molecular interactions in crystals may not fully correspond to those in the aqueous phase [35, 36]. In any case, the ambiguous chromatographic picture for Diogenes prompted us to try to evaluate its oligomeric state in a cellular model system.

To this end, we applied an OSER assay [23], which has become the *de facto* standard for assessing the monomerization of fluorescent proteins in cellulo [12, 37]. Since modern uses of the OSER assay often diverge from the original one, we performed the assessment in two different cell lines: HeLa for the widely-used simplified (OSER to non-OSER phenotypic) assessment and U2OS for the original (OSER : NE ratio) assessment. In HeLa, the analysis revealed ~ 87% whorl-free cells (*Fig. 2A*), which could be interpreted as a relatively high monomeric grade lying between the FusionRed and mKate2 scores published previously [12]. However, in addition to the obvious OSER-negative/positive cells, we observed a well-represented (~ 23%) cell fraction possessing diverse labeling features, such as small puncta, dots, local areas with increased brightness, which probably should not be attributed to a typical tubular ER phenotype (we labeled this population "mixed phenotype", see *Fig. 2B* for details). The aforementioned structures can be an indication of protein aggregation or its non-specific interaction with the intracellular environment, which could probably limit its efficiency in some circumstances. As per the original OSER analysis protocol in U2OS, the revealed mean OSER : NE ratio of Diogenes is 2.175, with the median value of 1.986 and standard deviation of 0.9299. Despite the relatively large standard deviation, both the mean OSER : NE ratio and the median allow us to consider Diogenes monomeric, with the monomericity borderline set at OSER : NE ⩽ 2.3 ± 0.6 [23]. Finally, we assembled several mammalian expression constructs for visual eval-

uation of the effectiveness of mKate2-K67R/R197H when working in fusions. For this testing, we selected targets (cytoskeleton proteins) whose visualization quality, according to our experience, noticeably depended on the oligomeric status of the tag (*Fig. 3*). We subjectively rated the labeling quality as very high.

pH stability of Diogenes. Next, we compared the stability of fluorescence intensity between Diogenes and its relatives, mKate2 and FusionRed, within the wide pH range of 3–11 (*Figs. S8* and *S9*). Over all, the protein exhibited a high pH stability, similar to that of mKate2, which is among the most pH-stable RFPs. Specifically, it could sustain a fluorescence level of ⩾ 80% of the maximum within the most physiologically and biochemically relevant pH range of 6.5–9.5. In the acidic range (pH 3–6), Diogenes showed a lower relative brightness than FusionRed but slightly surpassed mKate2 (their p*K*a values were determined to be 6.1, 5.76 and 6.16, respectively). Unlike both counterparts, Diogenes fluorescence decreases abruptly in the strongly alkaline pH range of 10–11, although this acidity level is not that biologically relevant.

Fluorescence lifetime. We then measured the fluorescence decay kinetics of the purified mKate2-K67R/R197H (Diogenes) in an aqueous solution using the time-correlated single photon counting approach and three different instrument setups (*Figs. S10–S12*). Importantly, the decay was shown to be monophasic in all the cases, with a lifetime value of ~ 2.2 ns. Surprisingly, in contrast to it and the FusionRed protein, which showed a biphasic fluorescence decay and a mean lifetime of ~ 1.6 ns with every setup used, the lifetime of the parental mKate2 was noticeably dependent on the measurement equipment. Thus, upon

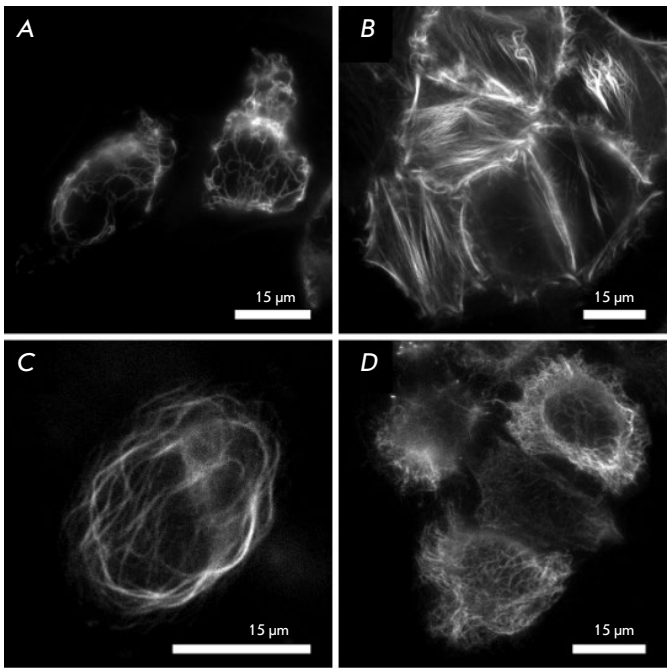


Fig. 3. Fluorescent labeling of intracellular structures with Diogenes in live HeLa Kyoto cells. (*A*) Vimentin–Diogenes; (*B*) lifeact–Diogenes; (*C*) ensconsin–Diogenes; and (*D*) Diogenes–cytokeratin; Scale bars are 15 μm

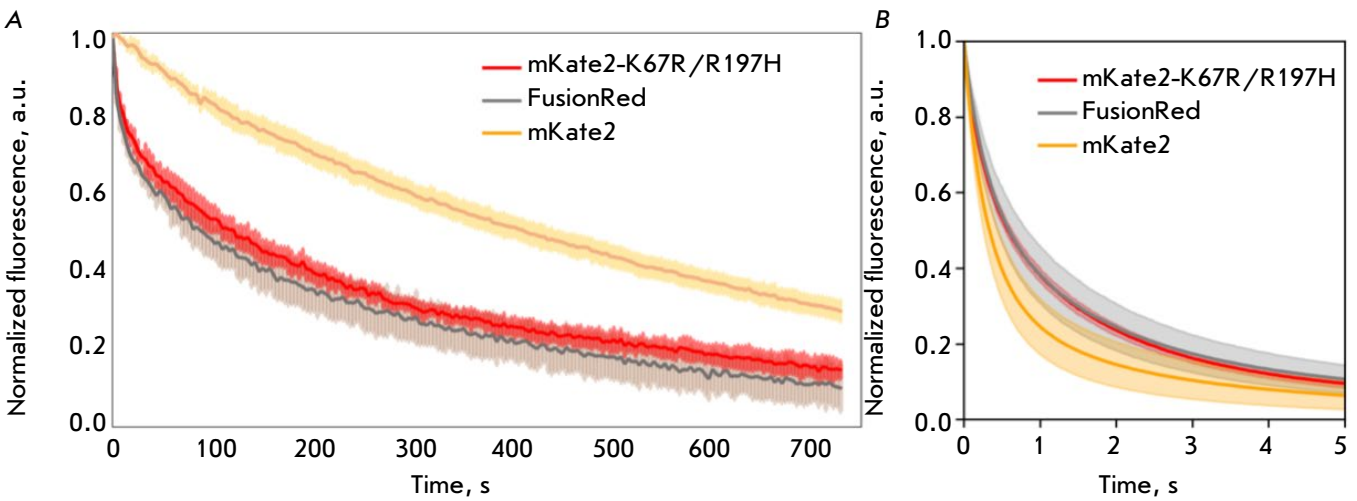


Fig. 4. The photobleaching kinetics of the red fluorescent proteins mKate2-K67R/R197H (Diogenes), mKate2, and FusionRed measured in an aqueous solution of the purified protein at an excitation power density of ~ 2 W/cm^2 (*A*) and live HeLa cells at ~ 1 kW/cm^2 (*B*). Solid lines indicate the mean fluorescence intensity during photobleaching. Transparent areas indicate the standard deviation (five protein-containing particles or 20 cells for each fluorescent protein)

excitation with a 450 nm picosecond laser (FWHM ~100 ps, 20 MHz) and a 590 nm nanosecond pulsed LED (FWHM ~1.5 ns, 20 MHz), its lifetime was 2.4 ns (*Figs. S10* and *S12*), while being only 2.05 ns upon excitation with a 590 nm femtosecond laser (FWHM ~150 fs, 80 MHz) (*Fig. S11*). The reasons for such flexibility remain unclear; they might be connected with some kind of excited-state processes known to occur in mKate2 and related proteins [30, 38]. Taking into account the excitation/emission wavelengths of Diogenes, mRuby [39] or mRuby2 [40] could be considered its close competitors in terms of fluorescence brightness/lifetime.

Photostability of Diogenes. High photostability is a desirable fluorophore property for both conventional fluorescence imaging and microscopy techniques with high spatial and temporal resolution [41]. Moreover, the photobleaching rate of a fluorescence protein may depend, non-linearly, on the excitation source power [12, 41]; this phenomenon can have a considerable impact when choosing a specific probe variant for a particular experiment. In this regard, we measured the photostability of Diogenes in two different model systems (*Fig. 4*). The photostability of purified mKate2-K67R/R197H measured in an aqueous environment (protein immobilized on microparticles) at a

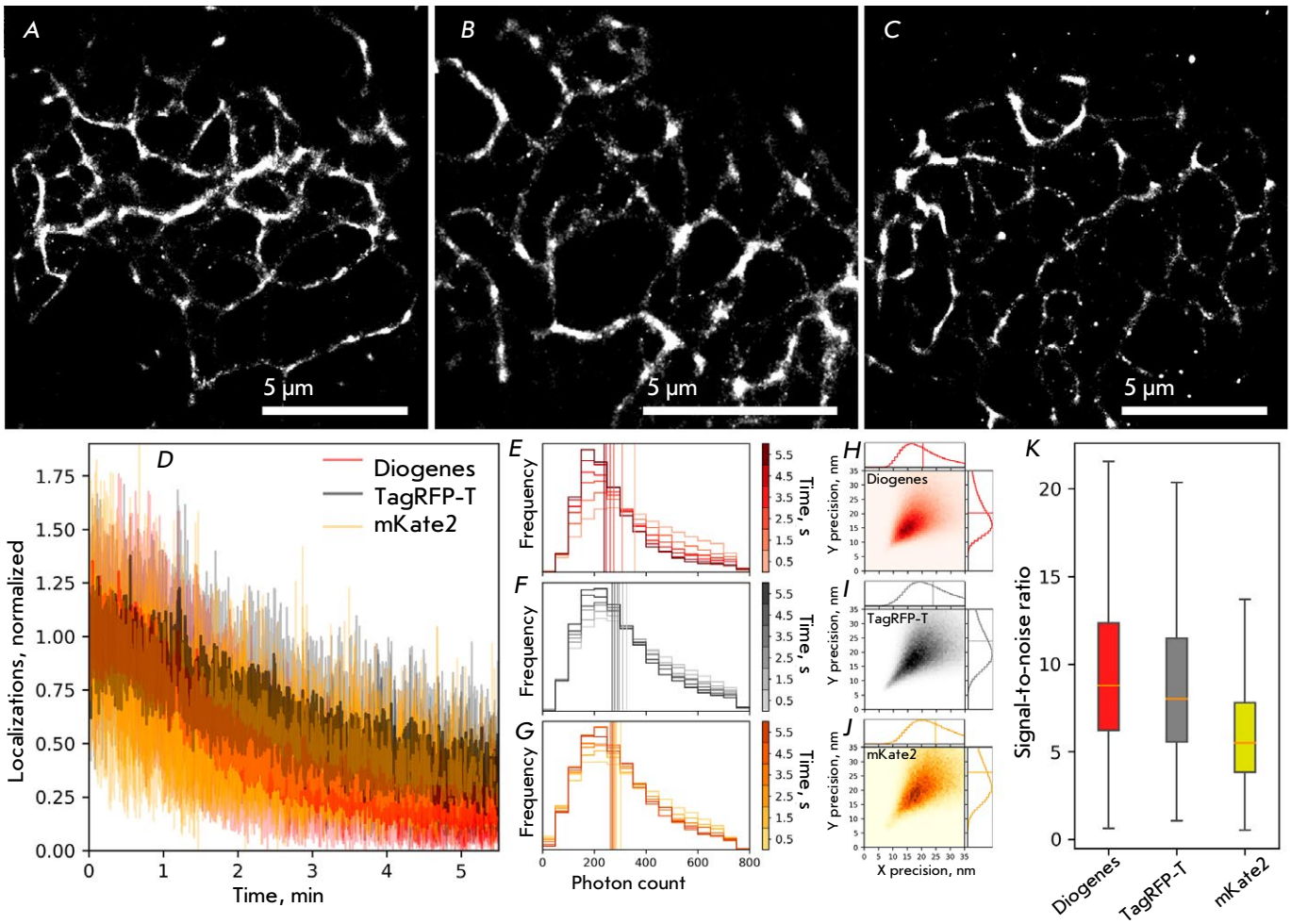


Fig. 5. Comparison of the live-cell super-resolution imaging performance of mKate2-K67R/R197H, TagRFP-T, and mKate2 as parts of vimentin fusion proteins in live HeLa cells under the following imaging conditions: 2 kW/cm^2 561 nm laser; 16.7 ms frame time; 20,000 frames. Super-resolution images of live HeLa cells transfected with vimentin–Diogenes, vimentin–TagRFP-T, and vimentin–mKate2, respectively (*A*), (*B*), (*C*); Scale bars are 5 μm. Stability of localization density of Diogenes, TagRFP-T, and mKate2 (*D*). The histogram of changes in the number of detected photons per single-molecule event over time of Diogenes, TagRFP-T, and mKate2, respectively; vertical lines represent the median values (*E*), (*F*), (*G*). 2D histograms of localization precision per single-molecule event of Diogenes, TagRFP-T, and mKate2, respectively (*H*), (*I*), (*J*); vertical lines on 1D histograms represent the median values. Signal-to-noise ratio of detected localizations (*K*); whiskers show the standard deviation, orange horizontal lines indicate median values

moderate power density typical of a widefield fluorescence microscope (~2 W/cm^2) was found to be slightly higher than that of FusionRed (bleaching $t_{1/2}$ 215 s vs. 165 s) and significantly lower than that of mKate2 ($t_{1/2}$ ~590 s, *Fig. 4A*). Surprisingly, in live HeLa cells upon high-intensity (~1 kW/cm^2) excitation, typical of SMLM techniques, the new RFP showed better performance than mKate2 (ca. twofold higher photostability, *Fig. 4B*) and approximately the same as FusionRed, which is much dimmer and was, therefore, expected to be more photostable.

Single-molecule behavior of mKate2-K67R/R197H

The increased photostability of Diogenes observed during imaging in a high excitation power mode, similar to that used in single-molecule microscopy techniques, prompted us to investigate the protein's behavior at the single-molecule level. Preliminary runs performed using dSTORM-like settings of the super-resolution fluorescence microscope on droplets of the purified protein revealed a pronounced stochastic blinking behavior of Diogenes (data not shown). Critically, red and far-red FPs, including variants such as mScarlet, mKate2, TagRFP, FusionRed, and FusionRed-MQ, although exhibiting blinking behavior [28, 42, 43], mostly fell short of green fluorescent proteins in terms of single-molecule performance, with only a few exceptions [42]. Therefore, evaluating the potential of new RFP variants for various SMLM techniques, where spontaneous fluorescence blinking can be utilized to refine the localization of labeled molecules, is important.

Here, we applied single-molecule localization microscopy (SMLM) to determine whether the Diogenes variant is capable of spontaneous blinking in cellulo and of visualizing intracellular structures with enhanced resolution. We chose the parental protein mKate2 and TagRFP-T, a protein known to have the strongest blinking pattern among the previously examined RFPs [42], as references. The single-molecule performance of Diogenes, TagRFP-T, and mKate2 was compared in a model system where these probes were fused with vimentin, transiently expressed in live HeLa cells, and monitored under conditions of super-resolution microscopy (*Fig. 5*).

Under the illumination of 2 kW/cm^2 at 561 nm light and with a 16.7 ms frame time, all the proteins blinked at a single molecule level, allowing for the reconstruction of the sub-diffraction image of vimentin fibers in live HeLa cells (*Fig. 5A–C*). Comparative analysis revealed that the stability of localization density was similar for all three proteins (*Fig. 5D*). Additionally, the difference in the median single-molecule brightness value was negligible among these proteins (*Fig. 5E–G*). The localization accuracy of Diogenes appeared to be slightly higher than that of mKate2 and TagRFP-T (20.5 nm vs. 24 nm and 25 nm, respectively, *Fig. 5H–J*), while the median value of the signal-to-noise ratio was ~1.5-fold higher for Diogenes than it was for mKate2, although the differences in the overall datasets for signal-to-noise values were insignificant (*Fig. 5K*). In conclusion, the performance of Diogenes in live-cell SMLM is on par or even slightly better than that of RFP, which was previously described as a promising probe for this microscopy modality [42]. However, it is still inferior in terms of the stability of localization density, molecular brightness, and localization precision to green fluorescent proteins capable of spontaneous blinking (e.g., mNeonGreen [44] and mBaoJin [45]). Intriguingly, additional 405 nm laser irradiation during imaging affected the density of localization of all three proteins (*Fig. S13*). Short violet laser flashes (illumination density of ~ 200 W/cm^2) significantly increased the number of recorded localizations of all three proteins. Although this experiment is not sufficient to draw a conclusion about the nature of this phenomenon, it is possible that short-wave illumination may induce additional maturation of the chromophore and/or switch the chromophore from the long-lived dark state to the fluorescent state (e.g., via *cis-trans* isomerization of the chromophore).

CONCLUSIONS

In this study, we systematically inspected the library of reciprocal mutants of the far-red fluorescent protein mKate2 and its daughter, the red FusionRed. We aimed to clarify the particular role of three residues in the chromophore environment (Arg/Lys-67, Cys/Ala-158, His/Arg-197) in determining the photophysical identities of these widely used genetically encoded probes.

One of the members of the constructed library, mKate2-K67R/R197H, named "Diogenes", exhibited a good combination of physicochemical and spectral properties, thus showing promise as a probe for conventional fluorescence microscopy techniques, as well as advanced imaging modalities of high spatial (SMLM) and temporal (FLIM) resolution. It inherits the advantages of both related proteins (FusionRed and mKate2). In particular, it possesses high fluorescence brightness, exhibits a monophasic fluorescence decay like mKate2, and shows good performance as a fusion tag similar to FusionRed. In terms of monomerization, Diogenes surpasses the parental mKate2 and possibly approaches the monomeric quality of FusionRed. The relatively high photostability of Diogenes (especially when normalized to the molecu-

lar brightness of the protein) under conditions of intense irradiation, as well as its remarkable ability for 405 nm illumination-induced photoactivation, which likely opens up possibilities for modulating its single-molecule behavior under live-cell multiphoton microscopy, are also worth noting. Our findings include indirect evidence that a smaller fraction of molecules trapped in long-lived transient dark states might be present in the population of Diogenes molecules (*Fig. S10*). Together with the absorption spectroscopy data (*Fig. 1A*), this may indicate a higher quality of chromophore maturation and steric adaptation inside the protein molecule compared to related RFPs.

It is important to note that, compared to its spectral analog, FusionRed, Diogenes carries a minimal number of mutations relative to the parental mKate2. Furthermore, the combination of substitutions (K67R/R197H) we found in the reciprocal library analysis was previously independently transferred from the bright but oligomerization-prone TagRFP protein to the dim monomer mKate2.5 to obtain FusionRed [10]. Like its relative FusionRed [18, 46], Diogenes can become a template for future semi-rational optimizations of RFPs, including those using high-throughput approaches. ●

This study was supported by the Russian Science Foundation (project No. 23-24-00011).

Supplementaries are available on the website https://doi.org/10.32607/actanaturae.27545.

REFERENCES

1. Cardarelli F. // Int. J. Mol. Sci. 2020. V. 21. № 11. P. 4164.
2. Rodriguez E.A., Campbell R.E., Lin J.Y., Lin M.Z., Miyawaki A., Palmer A.E., Shu X., Zhang J., Tsien R.Y. // Trends Biochem. Sci. 2017. V. 42. № 2. P. 111–129.
3. Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G., Markelov M.L., Lukyanov S.A. // Nat. Biotechnol. 1999. V. 17. № 10. P. 969–973.
4. Wiedenmann J., Schenk A., Röcker C., Girod A., Spindler K.-D., Nienhaus G.U. // Proc. Natl. Acad. Sci. U. S. A. 2002. V. 99. № 18. P. 11646–11651.
5. Schnitzler C.E., Keenan R.J., McCord R., Matysik A., Christianson L.M., Haddock S.H.D. // Mar. Biotechnol. N. Y. N. 2008. V. 10. № 3. P. 328–342.
6. Merzlyak E.M., Goedhart J., Shcherbo D., Bulina M.E., Shcheglov A.S., Fradkov A.F., Gaintzeva A., Lukyanov K.A., Lukyanov S., Gadella T.W.J., et al. // Nat. Methods. 2007. V. 4. № 7. P. 555–557.
7. Weissleder R. // Nat. Biotechnol. 2001. V. 19. № 4. P. 316–317.
8. Frangioni J.V. // Curr. Opin. Chem. Biol. 2003. V. 7. № 5. P. 626–634.
9. Smith A.M., Mancini M.C., Nie S. // Nat. Nanotechnol. 2009. V. 4. № 11. P. 710–711.
10. Shemiakina I.I., Ermakova G.V., Cranfill P.J., Baird M.A., Evans R.A., Souslova E.A., Staroverov D.B., Gorokhovatsky A.Y., Putintseva E.V., Gorodnicheva T.V., et al. // Nat. Commun. 2012. V. 3. P. 1204.
11. Costantini L.M., Baloban M., Markwardt M.L., Rizzo M.A., Guo F., Verkhusha V.V., Snapp E.L. // Nat. Commun. 2015. V. 6. P. 7670.
12. Cranfill P.J., Sell B.R., Baird M.A., Allen J.R., Lavagnino Z., de Gruiter H.M., Kremers G.-J., Davidson M.W., Ustione A., Piston D.W. // Nat. Methods. 2016. V. 13. № 7. P. 557–562.
13. Kost L.A., Nikitin E.S., Ivanova V.O., Sung U., Putintseva E.V., Chudakov D.M., Balaban P.M., Lukyanov K.A., Bogdanov A.M. // PloS One. 2017. V. 12. № 9. P. e0184225.
14. Shen Y., Dana H., Abdelfattah A.S., Patel R., Shea J., Molina R.S., Rawal B., Rancic V., Chang Y.-F., Wu L., et al. // BMC Biol. 2018. V. 16. № 1. P. 9.
15. Kost L.A., Ivanova V.O., Balaban P.M., Lukyanov K.A., Nikitin E.S., Bogdanov A.M. // Sensors. 2019. V. 19. № 13. P. 2982.
16. Yoon S., Pan Y., Shung K., Wang Y. // Sensors. 2020. V. 20. № 17. P. 4998.
17. Manna P., Hung S.-T., Mukherjee S., Friis P., Simpson D.M., Lo M.N., Palmer A.E., Jimenez R. // Integr. Biol. Quant. Biosci. Nano Macro. 2018. V. 10. № 9. P. 516–526.
18. Mukherjee S., Hung S.-T., Douglas N., Manna P., Thomas C., Ekrem A., Palmer A.E., Jimenez R. // Biochemistry. 2020. V. 59. № 39. P. 3669–3682.
19. Lambert T.J. // Nat. Methods. 2019. V. 16. № 4. P. 277–278.
20. Muslinkina L., Pletnev V.Z., Pletneva N.V., Ruchkin D.A., Kolesov D.V., Bogdanov A.M., Kost L.A., Rakitina T.V., Agapova Y.K., Shemyakina I.I., et al. // Int. J. Biol. Macromol. 2020. V. 155. P. 551–559.
21.Shcherbo D., Murphy C.S., Ermakova G.V., Solovieva E.A., Chepurnykh T.V., Shcheglov A.S., Verkhusha V.V., Pletnev V.Z., Hazelwood K.L., Roche P.M., et al. // Biochem. J. 2009. V. 418. № 3. P. 567–574.
22. García-Nafría J., Watson J.F., Greger I.H. // Sci. Rep. 2016. V. 6. P. 27459.
23. Costantini L.M., Fossati M., Francolini M., Snapp E.L. // Traffic Cph. Den. 2012. V. 13. № 5. P. 643–649.
24. Engler C., Gruetzner R., Kandzia R., Marillonnet S. // PloS One. 2009. V. 4. № 5. P. e5553.
25. Engler C., Kandzia R., Marillonnet S. // PloS One. 2008. V. 3. № 11. P. e3647.
26. Engler C., Marillonnet S. // Methods Mol. Biol. Clifton NJ. 2011. V. 729. P. 167–181.
27. Ermakova Y.G., Pak V.V., Bogdanova Y.A., Kotlobay A.A., Yampolsky I.V., Shokhina A.G., Panova A.S., Marygin R.A., Staroverov D.B., Bilan D.S., et al. // Chem. Commun. 2018. V. 54. № 23. P. 2898–2901.
28. Manna P., Jimenez R. // J. Phys. Chem. B. 2015. V. 119. № 15. P. 4944–4954.
29. Schindelin J., Arganda-Carreras I., Frise E., Kaynig V., Longair M., Pietzsch T., Preibisch S., Rueden C., Saalfeld S., Schmid B., et al. // Nat. Methods. 2012. V. 9. № 7. P. 676–682.
30. Protasova E.A., Mishin A.S., Lukyanov K.A., Maksimov E.G., Bogdanov A.M. // Photochem. Photobiol. Sci. Off. J.

Eur. Photochem. Assoc. Eur. Soc. Photobiol. 2021. V. 20. № 6. P. 791–803.
31. Subach F.V., Verkhusha V.V. // Chem. Rev. 2012. V. 112. № 7. P. 4308–4327.
32. Wachter R.M., Elsliger M.A., Kallio K., Hanson G.T., Remington S.J. // Struct. Lond. Engl. 1993. 1998. V. 6. № 10. P. 1267–1277.
33. Bindels D.S., Haarbosch L., van Weeren L., Postma M., Wiese K.E., Mastop M., Aumonier S., Gotthard G., Royant A., Hink M.A., et al. // Nat. Methods. 2017. V. 14. № 1. P. 53–56.
34. Pletnev S., Shcherbo D., Chudakov D.M., Pletneva N., Merzlyak E.M., Wlodawer A., Dauter Z., Pletnev V. // J. Biol. Chem. 2008. V. 283. № 43. P. 28980–28987.
35. Janin J., Rodier F. // Proteins. 1995. V. 23. № 4. P. 580–587.
36. Luo J., Liu Z., Guo Y., Li M. // Sci. Rep. 2015. V. 5. P. 14214.
37. Kim B.B., Wu H., Hao Y.A., Pan M., Chavarha M., Zhao Y., Westberg M., St-Pierre F., Wu J.C., Lin M.Z. // Sci. Rep. 2022. V. 12. № 1. P. 3678.
38. Kremers G.-J., Hazelwood K.L., Murphy C.S., Davidson M.W., Piston D.W. // Nat. Methods. 2009. V. 6. № 5. P. 355–358.
39. Kredel S., Oswald F., Nienhaus K., Deuschle K., Röcker C., Wolff M., Heilker R., Nienhaus G.U., Wiedenmann J. // PloS One. 2009. V. 4. № 2. P. e4391.
40. Lam A.J., St-Pierre F., Gong Y., Marshall J.D., Cranfill P.J., Baird M.A., McKeown M.R., Wiedenmann J., Davidson M.W., Schnitzer M.J., et al. // Nat. Methods. 2012. V. 9. № 10. P. 1005–1012.
41. Mamontova A.V., Grigoryev A.P., Tsarkova A.S., Lukyanov K.A., Bogdanov A.M. // Russ. J. Bioorganic Chem. 2017. V. 43. № 6. P. 625–633.
42. Klementieva N.V., Pavlikov A.I., Moiseev A.A., Bozhanova N.G., Mishina N.M., Lukyanov S.A., Zagaynova E.V., Lukyanov K.A., Mishin A.S. // Chem. Commun. Camb. Engl. 2017. V. 53. № 5. P. 949–951.
43. Mukherjee S., Thomas C., Wilson R., Simmerman E., Hung S.-T., Jimenez R. // Phys. Chem. Chem. Phys. PCCP. 2022. V. 24. № 23. P. 14310–14323.
44. Gavrikov A.S., Baranov M.S., Mishin A.S. // Biochem. Biophys. Res. Commun. 2020. V. 522. № 4. P. 852–854.
45. Zhang H., Lesnov G.D., Subach O.M., Zhang W., Kuzmicheva T.P., Vlaskina A.V., Samygina V.R., Chen L., Ye X., Nikolaeva A.Y., et al. // Nat. Methods. 2024. V. 21. № 4. P. 657–665.
46.Mukherjee S., Douglas N., Jimenez R. // J. Phys. Chem. Lett. 2024. V. 15. № 6. P. 1644–1651.

GENERAL RULES

Acta Naturae publishes experimental articles and reviews, as well as articles on topical issues, short reviews, and reports on the subjects of basic and applied life sciences and biotechnology.

The journal *Acta Naturae* is on the list of the leading periodicals of the Higher Attestation Commission of the Russian Ministry of Education and Science. The journal Acta Naturae is indexed in PubMed, Web of Science, Scopus and RCSI databases.

The editors of *Acta Naturae* ask of the authors that they follow certain guidelines listed below. Articles which fail to conform to these guidelines will be rejected without review. The editors will not consider articles whose results have already been published or are being considered by other publications.

The maximum length of a review, together with tables and references, cannot exceed 50 000 characters with spaces (approximately 30 pages, A4 format, 1.5 spacing, Times New Roman font, size 12) and cannot contain more than 16 figures.

Experimental articles should not exceed 30 000 symbols (approximately 15 pages in A4 format, including tables and references). They should contain no more than ten figures.

A short report must include the study's rationale, experimental material, and conclusions. A short report should not exceed 12 000 symbols (5–6 pages in A4 format including no more than 12 references). It should contain no more than three figures.

The manuscript and all necessary files should be uploaded to www.actanaturae.ru:

1) text in Word 2003 for Windows format;
2) the figures in TIFF format;
3) the text of the article and figures in one pdf file;
4) the article's title, the names and initials of the authors, the full name of the organizations, the abstract, keywords, abbreviations, figure captions, and Russian references should be translated to English;
5) the cover letter stating that the submitted manuscript has not been published elsewhere and is not under consideration for publication;
6) the license agreement (the agreement form can be downloaded from the website www.actanaturae.ru).

MANUSCRIPT FORMATTING

The manuscript should be formatted in the following manner:

- Article title. Bold font. The title should not be too long or too short and must be informative. The title should not exceed 100 characters. It should reflect the major result, the essence, and uniqueness of the work, names and initials of the authors.
- The corresponding author, who will also be working with the proofs, should be marked with a footnote *.
- Full name of the scientific organization and its departmental affiliation. If there are two or more scientific organizations involved, they should be linked by digital superscripts with the authors' names. Abstract. The structure of the abstract should be very clear and must reflect the following: it should introduce the reader to the main issue and describe the experimental approach, the possibility of practical use, and the possibility of further research in the field. The average length of an abstract is 20 lines (1 500 characters).
- Keywords (3 – 6). These should include the field of research, methods, experimental subject, and the specifics of the work. List of abbreviations.
- INTRODUCTION
- EXPERIMENTAL PROCEDURES
- RESULTS AND DISCUSSION
- CONCLUSION

  The organizations that funded the work should be listed at the end of this section with grant numbers in parenthesis.
- REFERENCES

  The in-text references should be in brackets, such as [1].

RECOMMENDATIONS ON THE TYPING AND FORMATTING OF THE TEXT

- We recommend the use of Microsoft Word 2003 for Windows text editing software.
- The Times New Roman font should be used. Standard font size is 12.
- The space between the lines is 1.5.
- Using more than one whole space between words is not recommended.
- We do not accept articles with automatic referencing; automatic word hyphenation; or automatic prohibition of hyphenation, listing, automatic indentation, etc.
- We recommend that tables be created using Word software options (Table → Insert Table) or MS Excel. Tables that were created manually (using lots of spaces without boxes) cannot be accepted.
- Initials and last names should always be separated by a whole space; for example, A. A. Ivanov.
- Throughout the text, all dates should appear in the "day.month.year" format, for example 02.05.1991, 26.12.1874, etc.
- There should be no periods after the title of the article, the authors' names, headings and subheadings, figure captions, units (s – second, g – gram, min – minute, h – hour, d – day, deg – degree).
- Periods should be used after footnotes (including those in tables), table comments, abstracts, and abbreviations (mon. – months, y. – years, m. temp. – melting temperature); however, they should not be used in subscripted indexes (T_m – melting temperature; $T_{p.t}$ – temperature of phase transition). One exception is mln – million, which should be used without a period.
- Decimal numbers should always contain a period and not a comma (0.25 and not 0,25).
- The hyphen ("-") is surrounded by two whole spaces, while the "minus," "interval," or "chemical bond" symbols do not require a space.

- The only symbol used for multiplication is "×"; the "×" symbol can only be used if it has a number to its right. The "·" symbol is used for denoting complex compounds in chemical formulas and also noncovalent complexes (such as DNA·RNA, etc.).
- Formulas must use the letter of the Latin and Greek alphabets.
- Latin genera and species' names should be in italics, while the taxa of higher orders should be in regular font.
- Gene names (except for yeast genes) should be italicized, while names of proteins should be in regular font.
- Names of nucleotides (A, T, G, C, U), amino acids (Arg, Ile, Val, etc.), and phosphonucleotides (ATP, AMP, etc.) should be written with Latin letters in regular font.
- Numeration of bases in nucleic acids and amino acid residues should not be hyphenated (T34, Ala89).
- When choosing units of measurement, SI units are to be used.
- Molecular mass should be in Daltons (Da, KDa, MDa).
- The number of nucleotide pairs should be abbreviated (bp, kbp).
- The number of amino acids should be abbreviated to aa.
- Biochemical terms, such as the names of enzymes, should conform to IUPAC standards.
- The number of term and name abbreviations in the text should be kept to a minimum.
- Repeating the same data in the text, tables, and graphs is not allowed.

GUIDENESS FOR ILLUSTRATIONS

- Figures should be supplied in separate files. Only TIFF is accepted.
- Figures should have a resolution of no less than 300 dpi for color and half-tone images and no less than 600 dpi.
- Files should not have any additional layers.

REVIEW AND PREPARATION OF THE MANUSCRIPT FOR PRINT AND PUBLICATION

Articles are published on a first-come, first-served basis. The members of the editorial board have the right to recommend the expedited publishing of articles which are deemed to be a priority and have received good reviews.

Articles which have been received by the editorial board are assessed by the board members and then sent for external review, if needed. The choice of reviewers is up to the editorial board. The manuscript is sent on to reviewers who are experts in this field of research, and the editorial board makes its decisions based on the reviews of these experts. The article may be accepted as is, sent back for improvements, or rejected.

The editorial board can decide to reject an article if it does not conform to the guidelines set above.

The return of an article to the authors for improvement does not mean that the article has been accepted for publication. After the revised text has been received, a decision is made by the editorial board. The author must return the improved text, together with the responses to all comments. The date of acceptance is the day on which the final version of the article was received by the publisher.

A revised manuscript must be sent back to the publisher a week after the authors have received the comments; if not, the article is considered a resubmission.

E-mail is used at all the stages of communication between the author, editors, publishers, and reviewers, so it is of vital importance that the authors monitor the address that they list in the article and inform the publisher of any changes in due time.

After the layout for the relevant issue of the journal is ready, the publisher sends out PDF files to the authors for a final review.

Changes other than simple corrections in the text, figures, or tables are not allowed at the final review stage. If this is necessary, the issue is resolved by the editorial board.

FORMAT OF REFERENCES

The journal uses a numeric reference system, which means that references are denoted as numbers in the text (in brackets) which refer to the number in the reference list.

Bibliographic descriptions of cited sources in the list of references should be drawn up in accordance with the requirements of the International Committee of Medical Journal Editors (ICMJE) and the AMA Manual of Style: A Guide for Authors and Editors (11th Edition).

Detailed rules for the design of links are presented here: https://actanaturae.ru/2075-8251/pages/view/references

The editorial board recommends the following resources to users of specialized bibliography software:

- EndNote — download the Acta Naturae.ens style file, view the JAMA reference style: Journal of the American Medical Association (AMA 11th edition);
- Zotero — view the American reference style Medical Association 11th edition (brackets);
- Mendeley — view the reference style of the American Medical Association.

Link design examples:

1. Hisakata R, Nishida S, Johnston A. An adaptable metric shapes perceptual space. *Curr Biol*. 2016;26(14):1911-1915. doi: 10.1016/j.cub.2016.05.047
2. Sambrook J, Russell DW. *Molecular Cloning: A Laboratory Manual*. 3rd ed. CSHL Press; 2001.
3. Hogue CWV. Structure databases. In: Baxevanis AD, Ouellette BFF, eds. *Bioinformatics*. 2nd ed. Life Sciences Series. Wiley-Interscience; 2001:83-109.

The following e-mail addresses can be used to contact the editorial staff: actanaturae@gmail.com, tel.: (495) 727-38-60.

SUPPLEMENTARY MATERIALS D. A. RUCHKIN, ET AL. "THE TWO KEY SUBSTITUTIONS IN THE CHROMOPHORE ENVIRONMENT OF mKate2 TO PRODUCE AN ENHANCED FUSIONRED-LIKE RED FLUORESCENT PROTEIN"

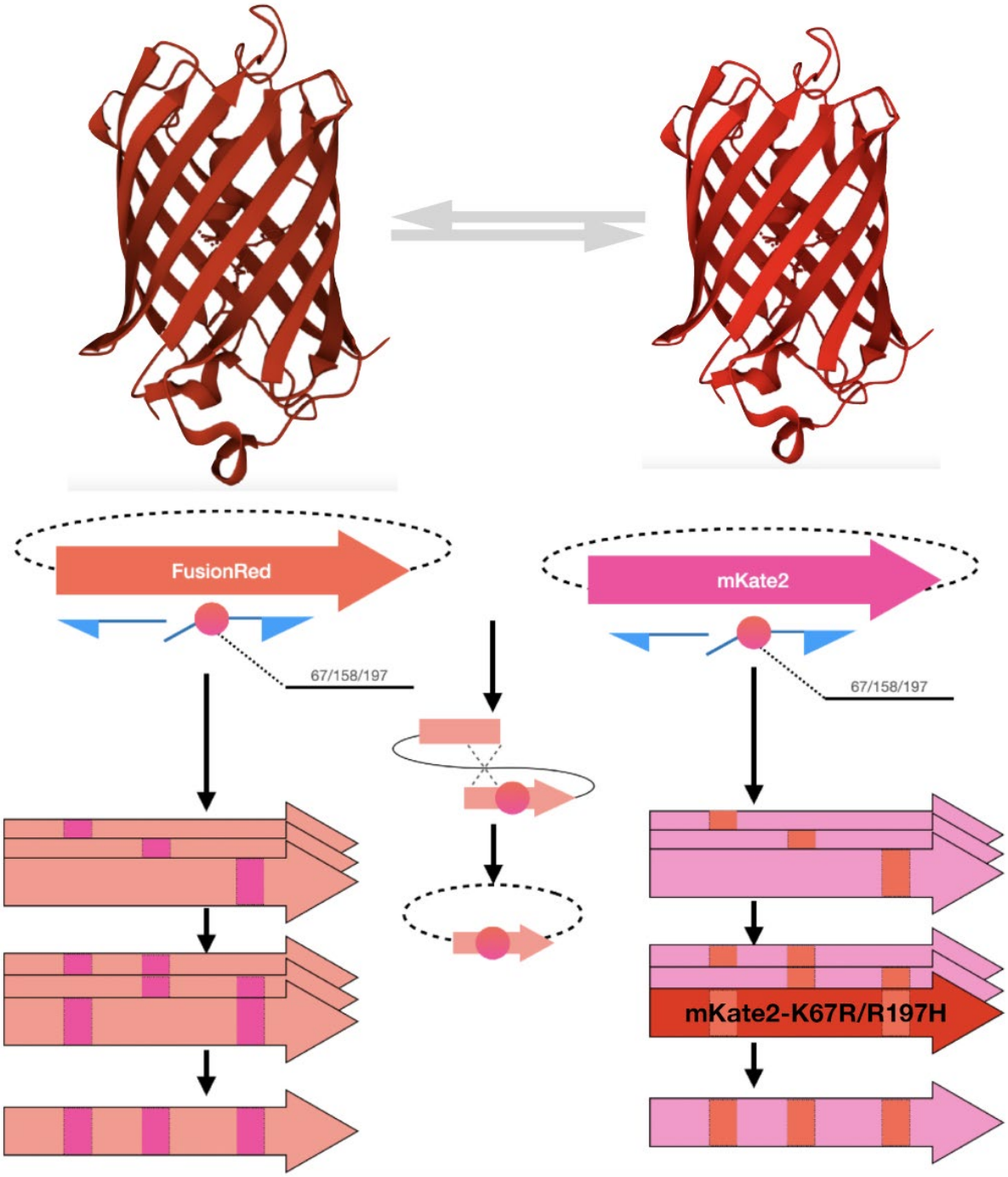


Fig. S1. Scheme showing how the library of point reciprocal mutants used in this study was engineered. The upper panel illustrates a general idea of 'exchanging' amino acid residues at positions 67/158/197 between the original FusionRed and mKate2 proteins. In the center of the scheme, there is a principle of IVA-cloning-based site-directed mutagenesis explained graphically (blue arrows are oligonucleotides; pink circles are DNA mismatches with substituted codons; dashed-lined ellipses depict vector backbone; the dotted cross shows recombination and ligation of the full-vector-PCR product). Aligned thick arrows on the left and right represent particular variants carrying single, double, and triple (single arrow) substitutions (mutated sites are shown as vertical bands of the "opposite" color)

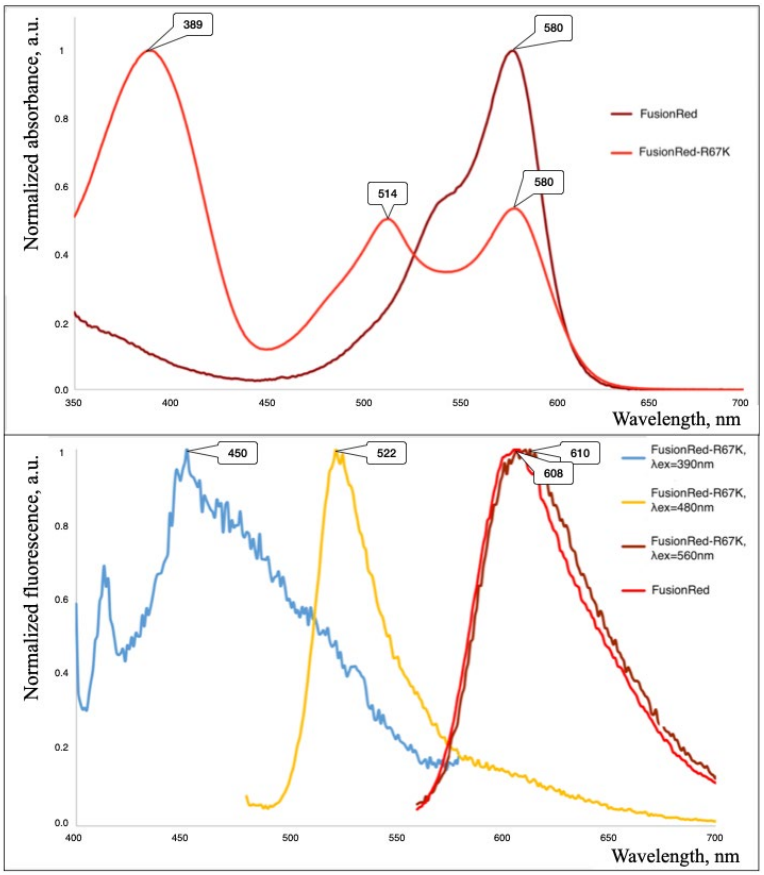


Fig. S2. Absorption (*A*) and fluorescence emission (*B*) spectra of FusionRed-R67K compared with those of its parent FusionRed. Wavelengths of the maxima of major bands are shown in the bubbles. In the fluorescence graph, excitation wavelengths used for recording the emission spectra are shown in the legend (the right upper corner)

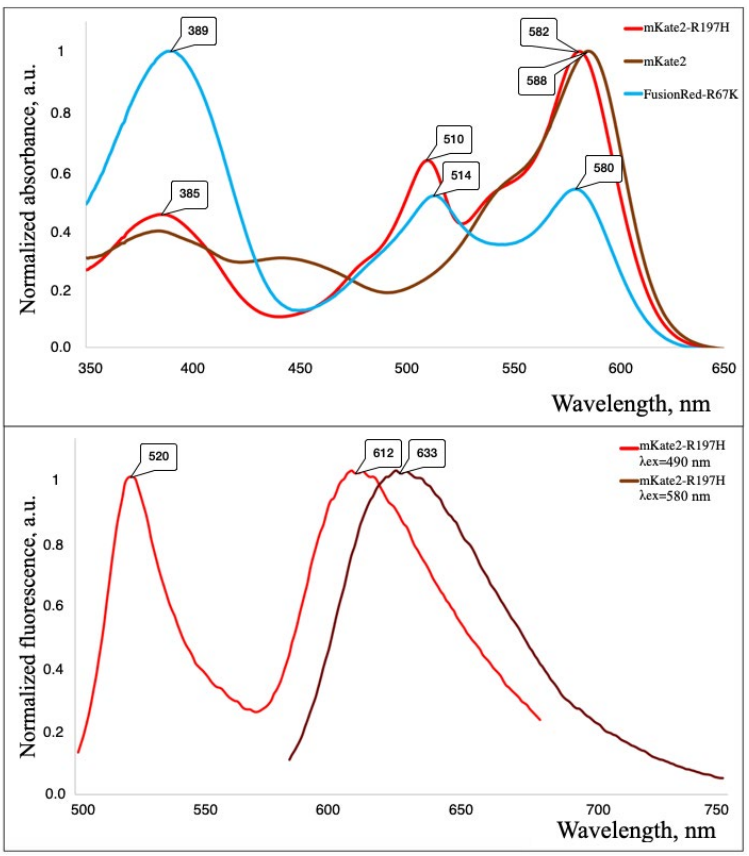


Fig. S3. Absorption (*A*) and fluorescence emission (*B*) spectra of mKate2-R197H compared with those of its parent mKate2. Wavelengths of the maxima of major bands are shown in the bubbles. In the absorption graph, the spectrum of FusionRed-R67K (solid blue line) is also added as a reference. In the fluorescence graph, excitation wavelengths used for recording the emission spectra are shown in the legend (the right upper corner)

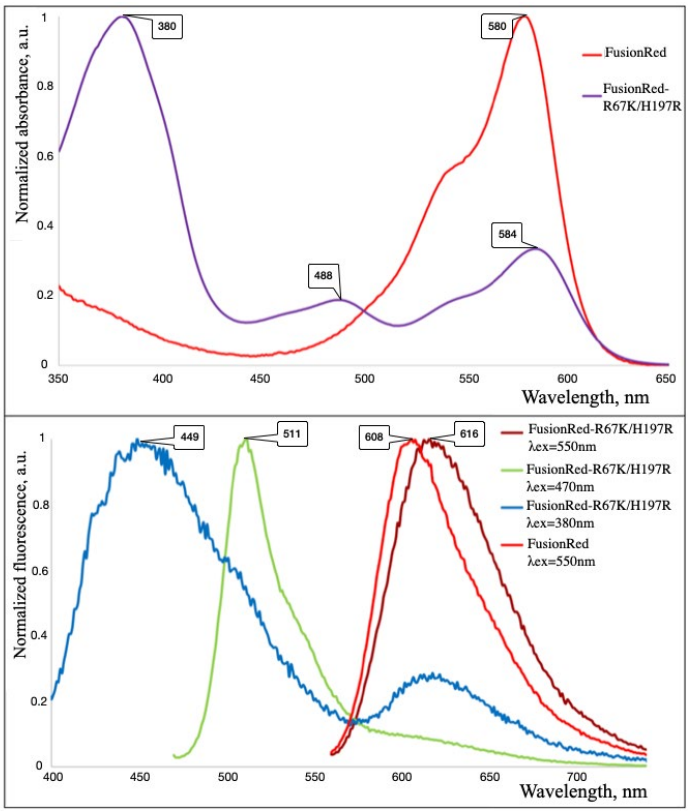


Fig. S4. Absorption (*A*) and fluorescence emission (*B*) spectra of FusionRed-R67K/H197R compared with those of its parent FusionRed. Wavelengths of the maxima of major bands are shown in the bubbles. In the fluorescence graph, excitation wavelengths used for the emission spectra recording are shown in the legend (the right upper corner)

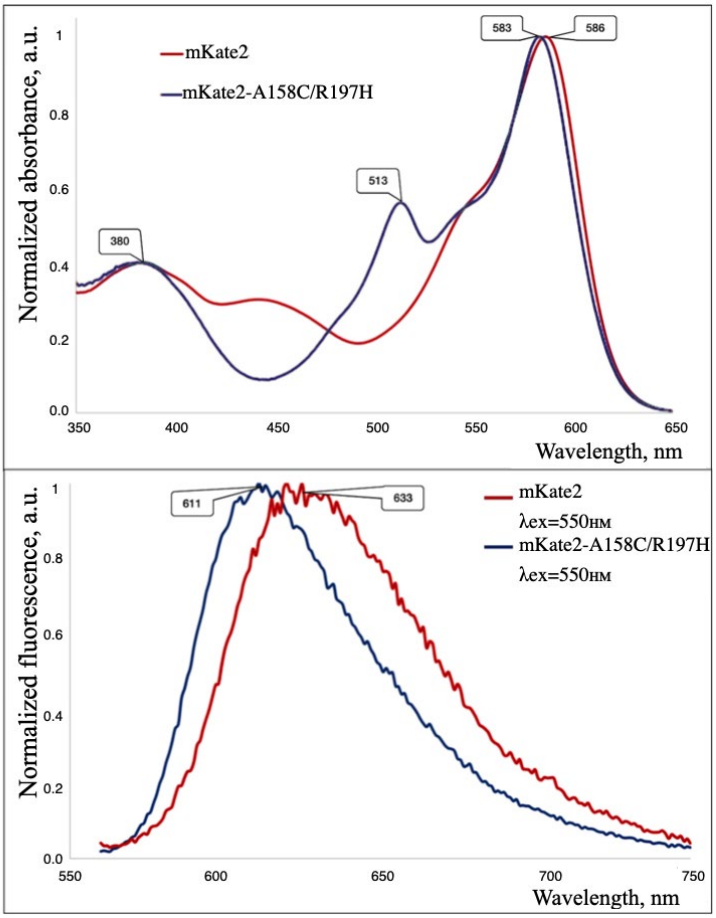


Fig. S5. Absorption (*A*) and fluorescence emission (*B*) spectra of mKate2-A158C/R197H compared with those of its parent mKate2. Wavelengths of the maxima of major bands are shown in the bubbles. In the fluorescence graph, excitation wavelengths used for recording the emission spectra are shown in the legend (the right upper corner)

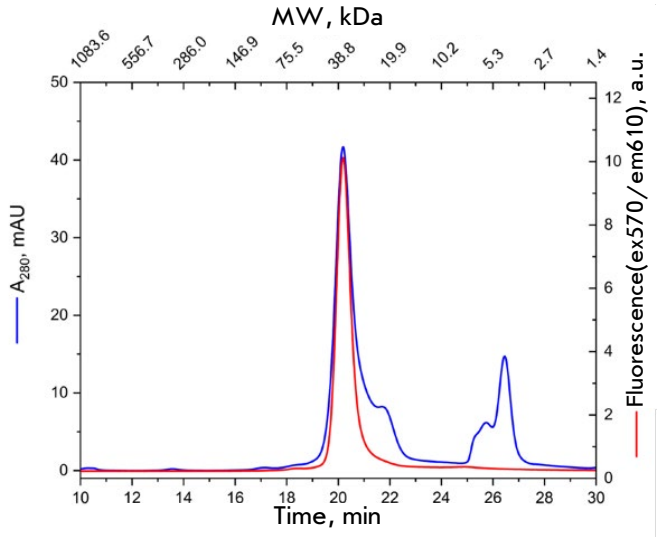


Fig. S6. Gel filtration chromatography of a purified mKate2-K67R/R197H (Diogenes) sample. Eluate was monitored using in-line absorbance and fluorescence detectors.

Table S1. Protein standards used to calibrate a gel filtration setup. The leftmost column represents the protein name; the central, molecular weight, kDa; the rightmost one, elution time while calibrating

Protein	Molecular mass, kDa	Elution time, min
CytochromeC, horse heart	12.4	23.354
Carboanhydrase, bovine erythrocytes	29	21.265
BSA	66	17.767
Alcohol dehydrogenase, yeast	150	15.995
β-amylase, sweet potato	200	15.253
Ferritin, horse spleen	450	12.723

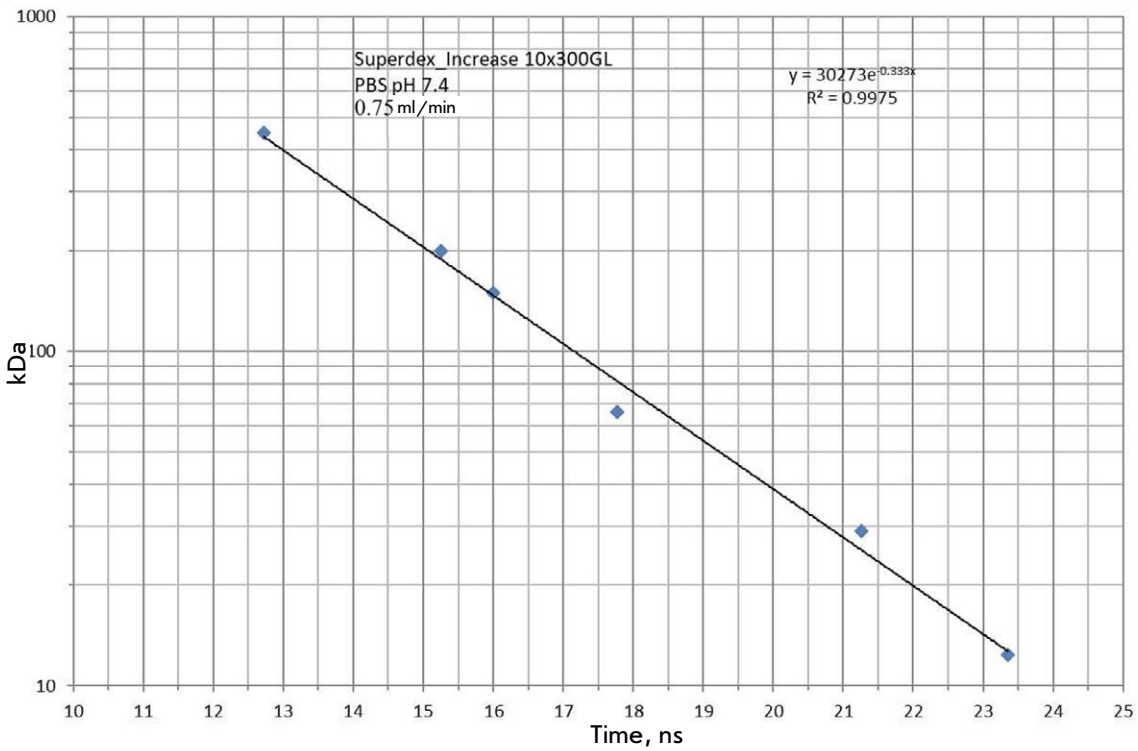


Fig. S7. Calibration plot used to determine molecular weights in the gel filtration/liquid chromatography setup (see the Materials and Methods section, main text). Blue squares show the protein standards used for calibration. Column type and equilibration/elution parameters, as well as the linear fitting results, are described on the graph

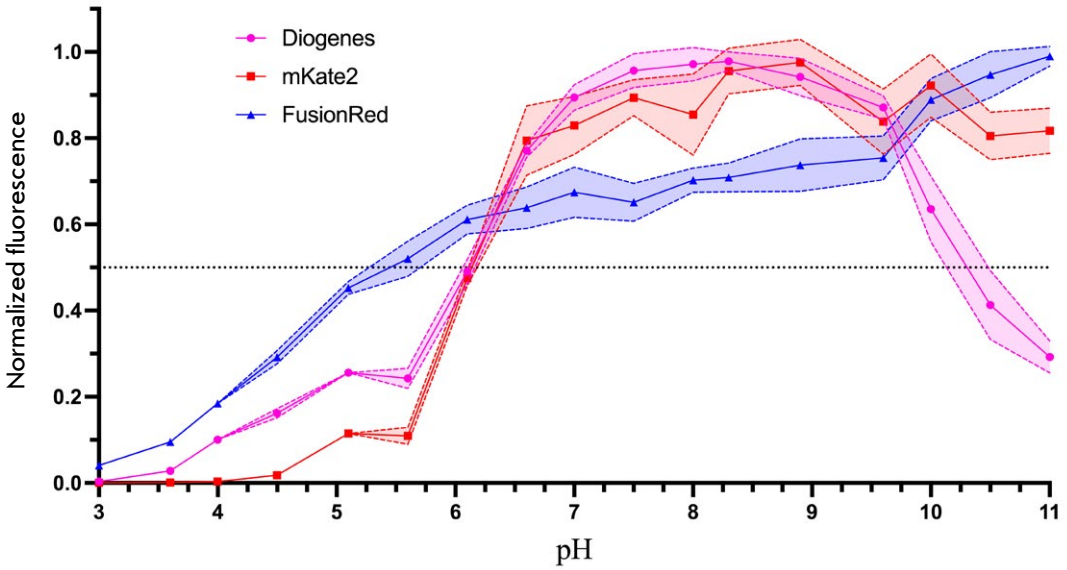


Fig. S8. Graph showing the fluorescence intensity dependence on pH (pH stability) measured for the purified mKate2-K67R/R197H (Diogenes), mKate2 and FusionRed proteins. The signal at each measurement point was normalized to a maximum signal value within the dataset

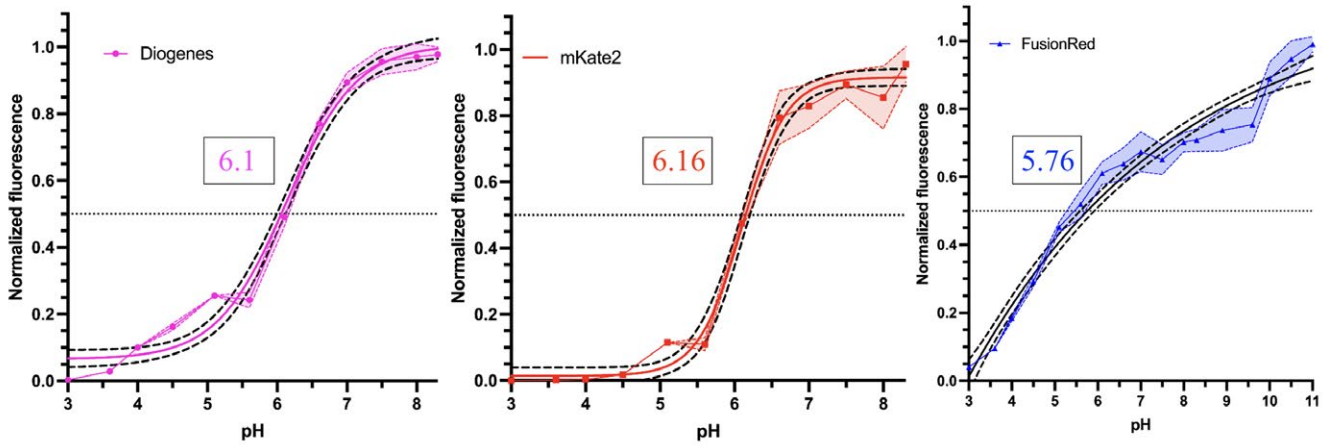


Fig. S9. Sigmoidal fits of the pH stability data for Diogenes, mKate2, and FusionRed. Colored graphs represent experimental data with standard deviations shown as semi-transparent areas ($n = 6$). The solid black line is the fitting curve with 95% confidence bands shown in black dashed lines. Fitting was performed using the GraphPad Prism10 package (the four-parameter logistic curve (4PL) fitting mode was used). The horizontal line stands for the half-maximum fluorescence level

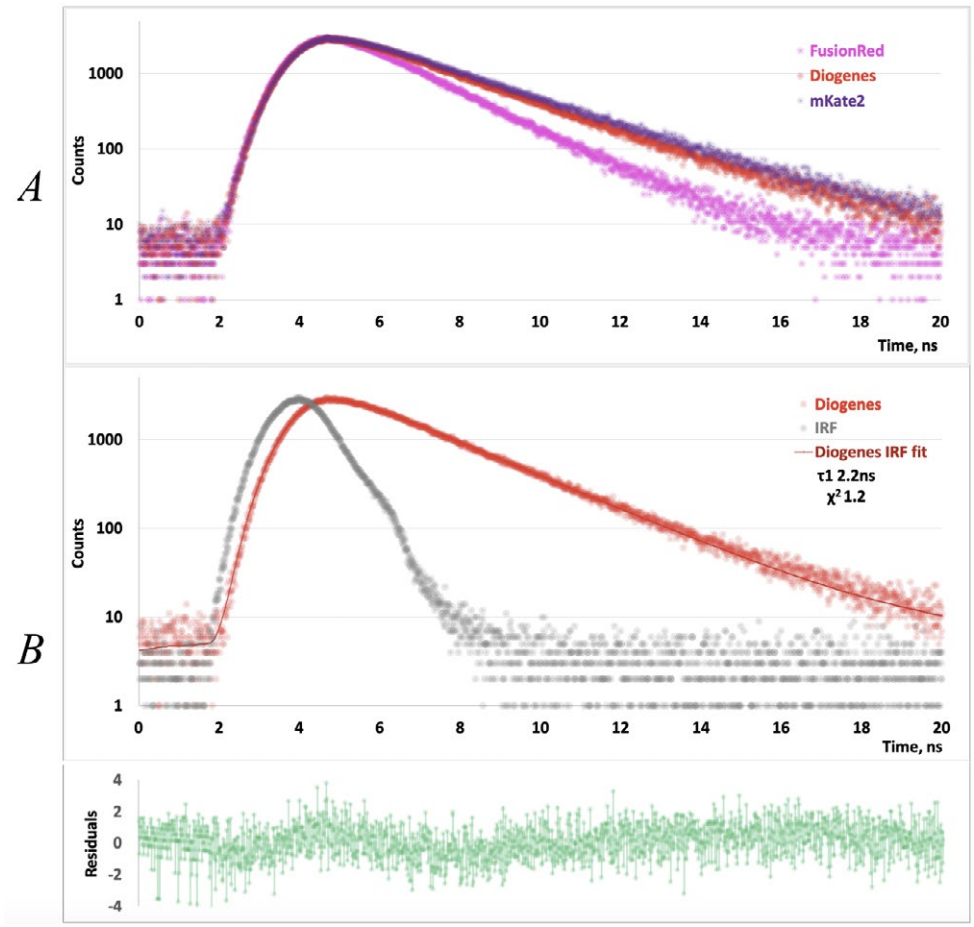


Fig. S10. The fluorescence decay kinetics recorded upon 590 nm single-photon excitation with the nanosecond pulsed LED driven at a repetition rate of 20 MHz. Comparison of the raw-data decay curves for FusionRed, mKate2, and Diogenes (*A*). Single-component exponential fitting of the Diogenes decay curve (*B*). Deconvolution with IRF was used for fitting. The measured instrument response (IRF) is shown in red

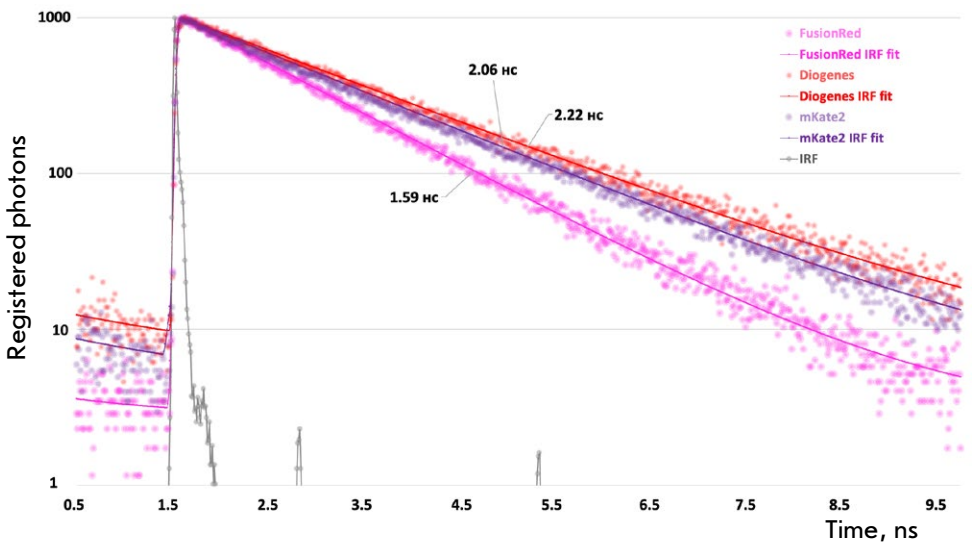


Fig. S11. The fluorescence decay kinetics recorded upon 590 nm single-photon excitation with the femtosecond laser driven at a repetition rate of 80 MHz. The raw-data and exponential fitting curves for FusionRed (biexponential fit), mKate2 and Diogenes (monoexponential fits) are shown. Mean lifetime values are inscribed next to the decay curves. Deconvolution with IRF was used for fitting. The measured instrument response (IRF) is shown in gray

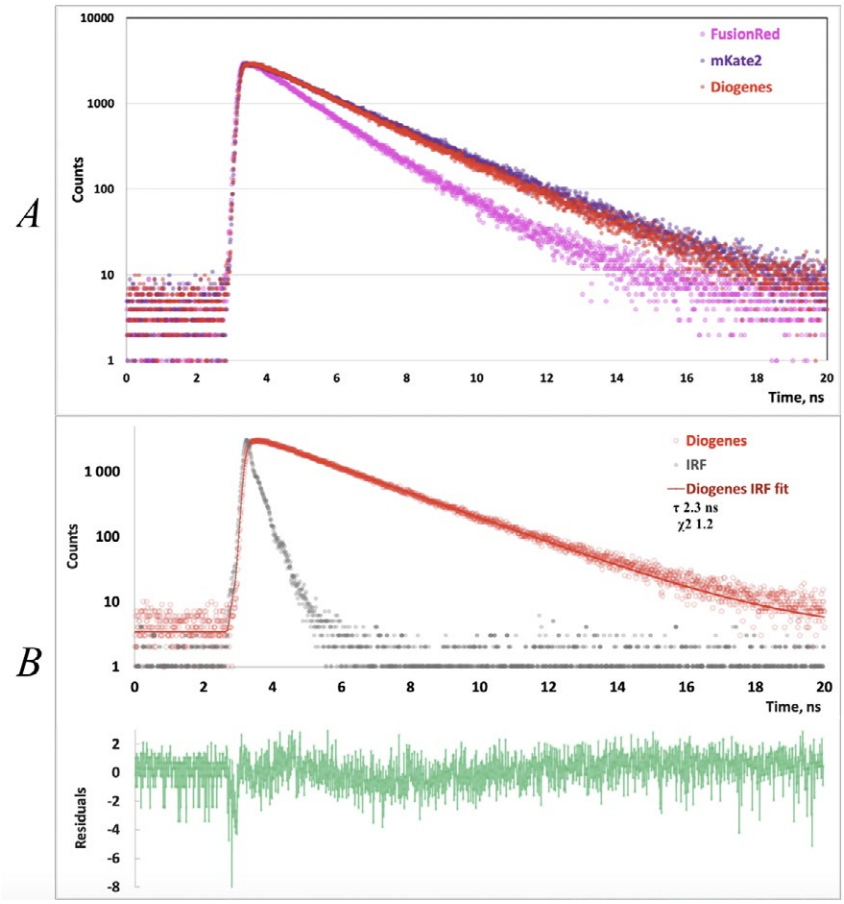


Fig. S12. The fluorescence decay kinetics recorded upon 450 nm single-photon excitation with the picosecond diode laser driven at a repetition rate of 20 MHz. Comparison of the raw-data decay curves for FusionRed, mKate2 and Diogenes (*A*). Single-component exponential fitting of the Diogenes decay curve (*B*). Deconvolution with IRF was used for fitting. The measured instrument response (IRF) is shown in gray

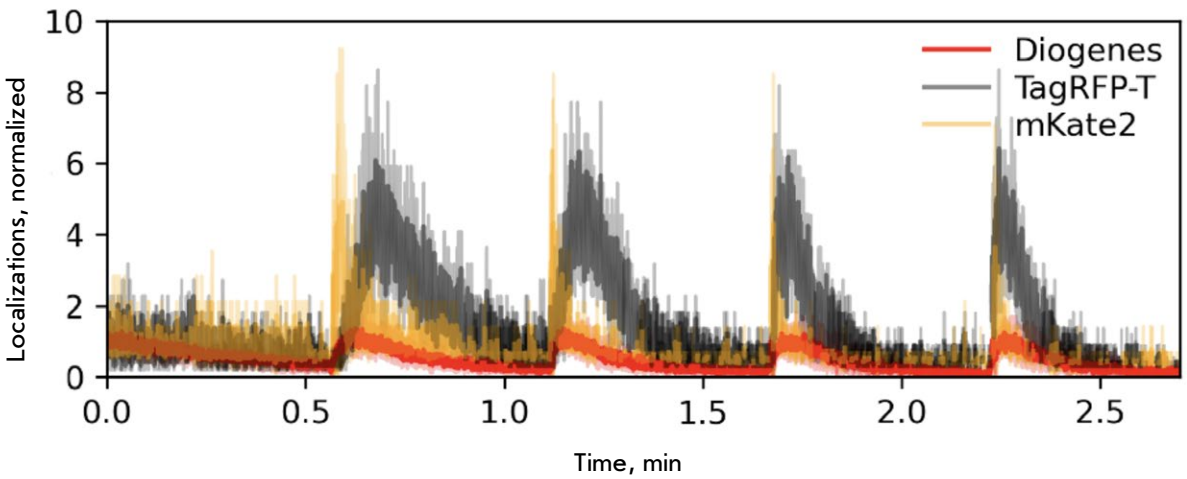


Fig. S13. The effect of 405 nm laser illumination on the localization density of Diogenes, TagRFP-T, and mKate2 as parts of vimentin fusion proteins in live HeLa cells during imaging under the following conditions: 2 kW/cm^2 561 nm laser, 16.7 ms frame time, 10,000 frames. Laser pulses were applied every 22 s with a duration of 405 nm laser illumination of 0.4 s and illumination density of ~215 W/cm^2

SUPPLEMENTARY MATERIALS K. A. IVANENKO, ET AL. "THE HYPOMETHYLATING AGENT 5-AZACITIDINE POTENTIATES THE EFFECT OF RAS AND Sp1 INHIBITORS IN NEUROBLASTOMA CELLS"

Table S1. Concentrations and manufacturers of the drugs used in the study

Name	Concentration	Manufacturer
5-Azacitidine	0–20 μM	Sigma-Aldrich
BI2536	10 nM	MedChemExpress
PD184352	10 nM	Selleckchem
Axitinib	0.5 μM	Sigma-Aldrich
Belinostat	0.1 μM	Selleckchem
Bortezomib	0.75 nM	Selleckchem
Volasertib	25 nM	Selleckchem
Vorinostat	1 μM	Selleckchem
Gefitinib	1 μM	Selleckchem
Dactolisib	50 nM	Selleckchem
Dexamethasone	25 μM	Selleckchem
Lonafarnib	5 μM	Selleckchem
Metformin	50 μM	Macklin
Mithramycin A	15 nM	Sigma-Aldrich
Palbociclib	0.1 μM	Macklin
Sorafenib	2.5 μM	Macklin
Staurosporine	10 nM	Macklin
Talazoparib	25 nM	Selleckchem
Entinostat	1 μM	MedChemExpress
Entrectinib	1 μM	Selleckchem

Table S1 lists the final concentrations of drugs used in the study.

Table S2. Characteristics of the dyes used in the study

Dye	Concentration	Staining time	Channel excitation wavelength, nm	Channel emission wavelength, nm
TMRE	0.1 μM	30 min	541–551	565–605
Tubulin Tracker™ Deep Red	0.33 μg/mL	30 min	590–650	662–738
LumiTracker® LysoGreen	0.1 μM	5 min	460–500	512–542
Hoechst-33342	1 μg/mL	30 min	325–375	435–485
HMRhoNox-M	2 μM	30 min	541–551	565–605
NucView® 488	2 μM	30 min	460–500	512–542
7-Aminoactinomycin	1 μg/mL	0 min	590–650	662–738

The cells were stained with dyes in the final concentrations specified in Table S2. The dyes were detected using a fluorescence microscope in the channels with specified excitation and emission wavelengths.

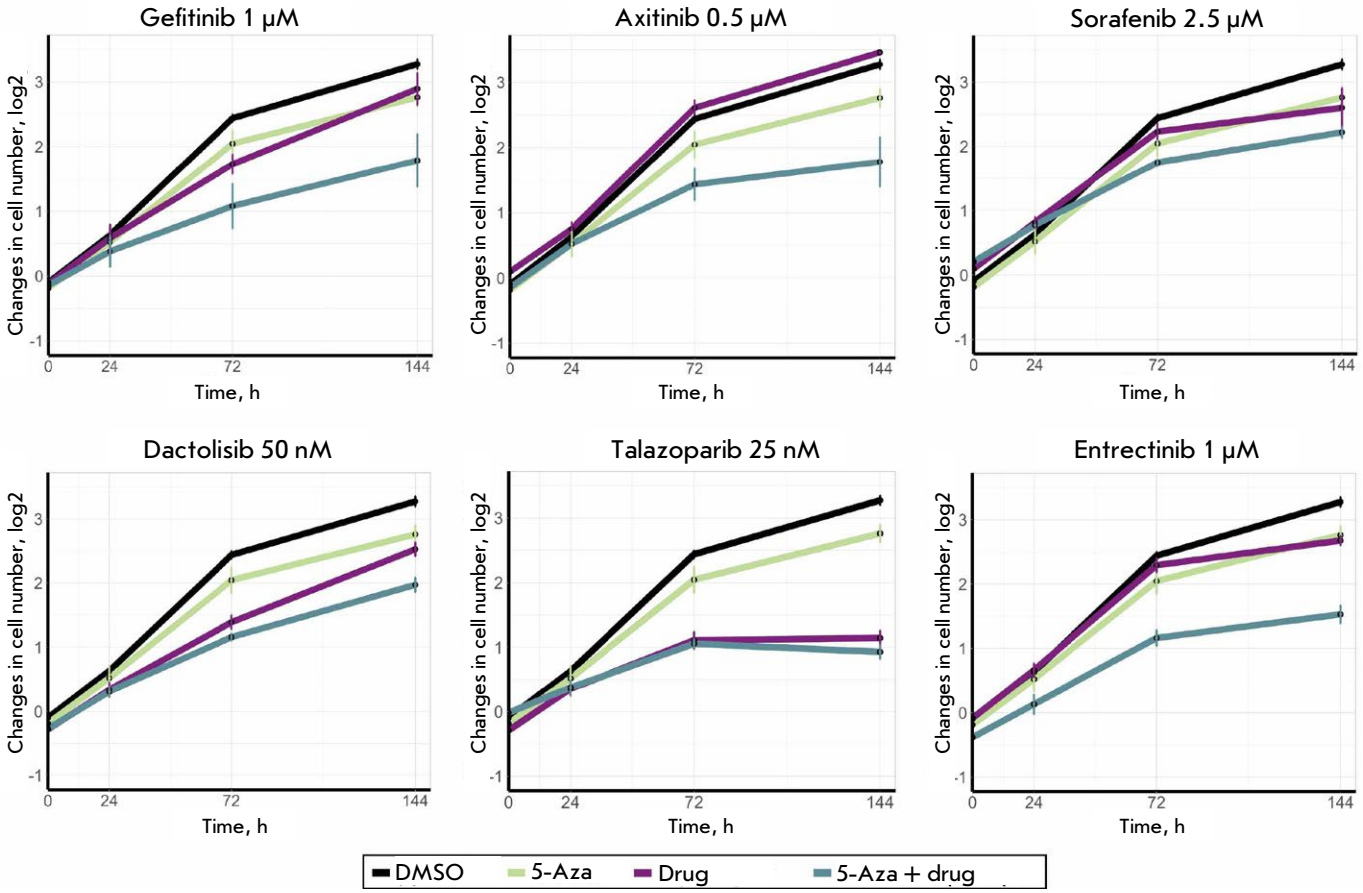


Fig. S1. The effectiveness of combinations of 5-azacitidine (5-Aza) with antitumor drugs for human neuroblastoma SH-SY5Y cells. The cells were simultaneously treated with 2.5 μM 5-Aza and an antitumor drug (drugs and their concentrations are shown in the figure) and co-incubated for 144 h. Cells incubated with dimethyl sulfoxide (DMSO) were used as the controls. The diagrams show the average value of three replicates; the standard deviation (SD) is provided

SUPPLEMENTARY MATERIALS I. S. ILNITSKIY, ET AL. "INTEGRATION OF HiMoRNA AND RNA-Chrom: VALIDATION OF THE FUNCTIONAL ROLE OF LONG NON-CODING RNAS IN THE EPIGENETIC REGULATION OF HUMAN GENES USING RNA-CHROMATIN INTERACTOME DATA"

Supplementary Table 1. lncRNA gene correspondence table between HiMoRNA and RNA-Chrom

https://himorna.fbras.ru/data/lncRNA_CorrespondenceTable.csv

Supplementary Table 2. Contingency table for calculating right-tailed and left-tailed Fisher's test, for example, for the pair lncRNA PVT1 – histone mark H3K27ac peaks

	Total number of "-" peaks H3K27ac	Total number of "+" peaks H3K27ac
Total number of "+" and "-" peaks supported by PVT1 contacts	Number of supported by contacts PVT1 "-" peaks H3K27ac	Number of supported by contacts PVT1 "+" peaks H3K27ac
Total number of "+" and "-" peaks not supported by PVT1 contacts	Number of unsupported PVT1 "-" peaks H3K27ac	Number of unsupported PVT1 "+" peaks H3K27ac

Supplementary Table 3. Step-by-step analysis of lncRNA MIR31HG using the integrated HiMoRNA and RNA-Chrom databases

Web resource	Step	Current page	Actions	Result
HiMoRNA	1	Homepage	Click "Search page" or "Search" button	Search page
	2	Search page	- add "MIR31HG" in the corresponding field - select the histone modifications of interest (H3K4me1, H3K27ac) - add the genomic coordinates of the extended *GLI2* promoter region (chr2:120725622-120992653) - click the "Search" button	Search results page
	Optional	Search results page	- click on the histone modification	Histone mark page
			- click on the lncRNA hyperlink	lncRNA page
			- click on the gene hyperlink	Histone modification peak associated gene page
			- click on the correlation coefficient hyperlink	Correlation coefficient page
			- click on the "Download" button	Download the table with the search results
	3	Search results page	- select the peak of interest using the switch to the left of the table - click on the "Go to RNA-Chrom DB" button, then on "MIR31HG RNA contacts with the locus (peak_473655)" and select the required peak coordinate extension (for example, 25000)	Graphical summary of lncRNA MIR31HG contacts with chromatin in the extended peak region
RNA-Chrom	4	Graphic summary page	- select the experiment of interest using the flag (Exp. ID: 9) - click on the button "ALL TARGET GENES"	Table with all genes from the extended peak region that MIR31HG interacts with or does not interact with

Web resource	Step	Current page	Actions	Result
HiMoRNA	5	Search page	- add “MIR31HG” to the corresponding field - select the histone modifications of interest (H3K4me1, H3K27ac) - add the names of 56 genes from the “Hedgehog signaling pathway” category to the “Gene/Gene ID (optional)” field – click the “Search” button	Search results page
	6	Search results page	- select peak H3K27ac_963553 (chr9:95446174-95452554) - click on the button “Go to RNA-Chrom DB”, then on “MIR31HG RNA contacts with the locus (peak_963553)” and select the required extension of peak coordinates (for example, 25000)	Graphical summary of lncRNA MIR31HG contacts with chromatin in the extended peak region
RNA-Chrom	7	Graphic summary page	- select the experiment of interest using the flag (Exp. ID: 5) - click on the button “ALL TARGET GENES”	Table with all genes from the extended peak region with which MIR31HG interacts or not, among which we find PTCH1

Supplementary Table 4. Genes from the “Hedgehog Signaling Pathway” obtained from the KEGG Pathway database

Original Alias	Converted Alias	Gene Name	Description	Namespace
KEGG:04340	ENSG00000010282	HHATL	hedgehog acyltransferase like [Source:HGNC Symbol;Acc:HGNC:13242]	KEGG
KEGG:04340	ENSG00000036257	CUL3	cullin 3 [Source:HGNC Symbol;Acc:HGNC:2553]	KEGG
KEGG:04340	ENSG00000054392	HHAT	hedgehog acyltransferase [Source:HGNC Symbol;Acc:HGNC:18270]	KEGG
KEGG:04340	ENSG00000055130	CUL1	cullin 1 [Source:HGNC Symbol;Acc:HGNC:2551]	KEGG
KEGG:04340	ENSG00000064309	CDON	cell adhesion associated, oncogene regulated [Source:HGNC Symbol;Acc:HGNC:17104]	KEGG
KEGG:04340	ENSG00000072062	PRKACA	protein kinase cAMP-activated catalytic subunit alpha [Source:HGNC Symbol;Acc:HGNC:9380]	KEGG
KEGG:04340	ENSG00000072803	FBXW11	F-box and WD repeat domain containing 11 [Source:HGNC Symbol;Acc:HGNC:13607]	KEGG
KEGG:04340	ENSG00000072840	EVC	EvC ciliary complex subunit 1 [Source:HGNC Symbol;Acc:HGNC:3497]	KEGG
KEGG:04340	ENSG00000074047	GLI2	GLI family zinc finger 2 [Source:HGNC Symbol;Acc:HGNC:4318]	KEGG
KEGG:04340	ENSG00000081479	LRP2	LDL receptor related protein 2 [Source:HGNC Symbol;Acc:HGNC:6694]	KEGG
KEGG:04340	ENSG00000082701	GSK3B	glycogen synthase kinase 3 beta [Source:HGNC Symbol;Acc:HGNC:4617]	KEGG
KEGG:04340	ENSG00000100077	GRK3	G protein-coupled receptor kinase 3 [Source:HGNC Symbol;Acc:HGNC:290]	KEGG
KEGG:04340	ENSG00000102858	MGRN1	mahogunin ring finger 1 [Source:HGNC Symbol;Acc:HGNC:20254]	KEGG

Original Alias	Converted Alias	Gene Name	Description	Namespace
KEGG:04340	ENSG00000105429	MEGF8	multiple EGF like domains 8 [Source:HGNC Symbol;Acc:HGNC:3233]	KEGG
KEGG:04340	ENSG00000106012	IQCE	IQ motif containing E [Source:HGNC Symbol;Acc:HGNC:29171]	KEGG
KEGG:04340	ENSG00000106571	GLI3	GLI family zinc finger 3 [Source:HGNC Symbol;Acc:HGNC:4319]	KEGG
KEGG:04340	ENSG00000107882	SUFU	SUFU negative regulator of hedgehog signaling [Source:HGNC Symbol;Acc:HGNC:16466]	KEGG
KEGG:04340	ENSG00000108854	SMURF2	SMAD specific E3 ubiquitin protein ligase 2 [Source:HGNC Symbol;Acc:HGNC:16809]	KEGG
KEGG:04340	ENSG00000110092	CCND1	cyclin D1 [Source:HGNC Symbol;Acc:HGNC:1582]	KEGG
KEGG:04340	ENSG00000111087	GLI1	GLI family zinc finger 1 [Source:HGNC Symbol;Acc:HGNC:4317]	KEGG
KEGG:04340	ENSG00000113712	CSNK1A1	casein kinase 1 alpha 1 [Source:HGNC Symbol;Acc:HGNC:2451]	KEGG
KEGG:04340	ENSG00000117425	PTCH2	patched 2 [Source:HGNC Symbol;Acc:HGNC:9586]	KEGG
KEGG:04340	ENSG00000118971	CCND2	cyclin D2 [Source:HGNC Symbol;Acc:HGNC:1583]	KEGG
KEGG:04340	ENSG00000121067	SPOP	speckle type BTB/POZ protein [Source:HGNC Symbol;Acc:HGNC:11254]	KEGG
KEGG:04340	ENSG00000128602	SMO	smoothened, frizzled class receptor [Source:HGNC Symbol;Acc:HGNC:11119]	KEGG
KEGG:04340	ENSG00000131437	KIF3A	kinesin family member 3A [Source:HGNC Symbol;Acc:HGNC:6319]	KEGG
KEGG:04340	ENSG00000133275	CSNK1G2	casein kinase 1 gamma 2 [Source:HGNC Symbol;Acc:HGNC:2455]	KEGG
KEGG:04340	ENSG00000137486	ARRB1	arrestin beta 1 [Source:HGNC Symbol;Acc:HGNC:711]	KEGG
KEGG:04340	ENSG00000139549	DHH	desert hedgehog signaling molecule [Source:HGNC Symbol;Acc:HGNC:2865]	KEGG
KEGG:04340	ENSG00000141480	ARRB2	arrestin beta 2 [Source:HGNC Symbol;Acc:HGNC:712]	KEGG
KEGG:04340	ENSG00000141551	CSNK1D	casein kinase 1 delta [Source:HGNC Symbol;Acc:HGNC:2452]	KEGG
KEGG:04340	ENSG00000142875	PRKACB	protein kinase cAMP-activated catalytic subunit beta [Source:HGNC Symbol;Acc:HGNC:9381]	KEGG
KEGG:04340	ENSG00000143147	GPR161	G protein-coupled receptor 161 [Source:HGNC Symbol;Acc:HGNC:23694]	KEGG
KEGG:04340	ENSG00000144228	SPOPL	speckle type BTB/POZ protein like [Source:HGNC Symbol;Acc:HGNC:27934]	KEGG
KEGG:04340	ENSG00000144857	BOC	BOC cell adhesion associated, oncogene regulated [Source:HGNC Symbol;Acc:HGNC:17173]	KEGG

Original Alias	Converted Alias	Gene Name	Description	Namespace
KEGG:04340	ENSG00000151292	CSNK1G3	casein kinase 1 gamma 3 [Source:HGNC Symbol;Acc:HGNC:2456]	KEGG
KEGG:04340	ENSG00000154309	DISP1	dispatched RND transporter family member 1 [Source:HGNC Symbol;Acc:HGNC:19711]	KEGG
KEGG:04340	ENSG00000163501	IHH	Indian hedgehog signaling molecule [Source:HGNC Symbol;Acc:HGNC:5956]	KEGG
KEGG:04340	ENSG00000164161	HHIP	hedgehog interacting protein [Source:HGNC Symbol;Acc:HGNC:14866]	KEGG
KEGG:04340	ENSG00000164690	SHH	sonic hedgehog signaling molecule [Source:HGNC Symbol;Acc:HGNC:10848]	KEGG
KEGG:04340	ENSG00000165059	PRKACG	protein kinase cAMP-activated catalytic subunit gamma [Source:HGNC Symbol;Acc:HGNC:9382]	KEGG
KEGG:04340	ENSG00000166167	BTRC	beta-transducin repeat containing E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:1144]	KEGG
KEGG:04340	ENSG00000166813	KIF7	kinesin family member 7 [Source:HGNC Symbol;Acc:HGNC:30497]	KEGG
KEGG:04340	ENSG00000169118	CSNK1G1	casein kinase 1 gamma 1 [Source:HGNC Symbol;Acc:HGNC:2454]	KEGG
KEGG:04340	ENSG00000171791	BCL2	BCL2 apoptosis regulator [Source:HGNC Symbol;Acc:HGNC:990]	KEGG
KEGG:04340	ENSG00000173020	GRK2	G protein-coupled receptor kinase 2 [Source:HGNC Symbol;Acc:HGNC:289]	KEGG
KEGG:04340	ENSG00000173040	EVC2	EvC ciliary complex subunit 2 [Source:HGNC Symbol;Acc:HGNC:19747]	KEGG
KEGG:04340	ENSG00000175356	SCUBE2	signal peptide, CUB domain and EGF like domain containing 2 [Source:HGNC Symbol;Acc:HGNC:30425]	KEGG
KEGG:04340	ENSG00000180138	CSNK1A1L	casein kinase 1 alpha 1 like [Source:HGNC Symbol;Acc:HGNC:20289]	KEGG
KEGG:04340	ENSG00000180447	GAS1	growth arrest specific 1 [Source:HGNC Symbol;Acc:HGNC:4165]	KEGG
KEGG:04340	ENSG00000185716	MOSMO	modulator of smoothened [Source:HGNC Symbol;Acc:HGNC:27087]	KEGG
KEGG:04340	ENSG00000185920	PTCH1	patched 1 [Source:HGNC Symbol;Acc:HGNC:9585]	KEGG
KEGG:04340	ENSG00000198742	SMURF1	SMAD specific E3 ubiquitin protein ligase 1 [Source:HGNC Symbol;Acc:HGNC:16807]	KEGG
KEGG:04340	ENSG00000203965	EFCAB7	EF-hand calcium binding domain 7 [Source:HGNC Symbol;Acc:HGNC:29379]	KEGG
KEGG:04340	ENSG00000213923	CSNK1E	casein kinase 1 epsilon [Source:HGNC Symbol;Acc:HGNC:2453]	KEGG
KEGG:04340	ENSG00000283900	TPTEP2-CSNK1E	TPTEP2-CSNK1E readthrough [Source:HGNC Symbol;Acc:HGNC:53829]	KEGG

Supplementary Table 5. Step-by-step analysis of lncRNA PVT1 using integrated databases HiMoRNA and RNA-Chrom

Web resource	Step	Current page	Actions	Result
HiMoRNA	1	Homepage	- click "Search page" or "Search" button	Search page
	2	Search page	- add "PVT1" in the corresponding field - select all histone modifications - add "LATS2" to the "Gene/Gene ID" field - click on the "Search" button	Search results page
	3	Search results page	- select the peak of interest using the switch to the left of the table (for example, peak_169403) - click on the "Go to RNA-Chrom DB" button, then on "PVT1 RNA contacts with the locus (peak_169403)" and select the required peak coordinate extension (for example, 25000)	Graphical summary of lncRNA PVT1 contacts with chromatin in the extended peak region
RNA-Chrom	4	Graphic summary page	- select the experiment of interest using the flag (Exp.IDs: 8, 10) - click on the button "VIEW IN GENOME BROWSER"	UCSC Genome Browser

SUPPLEMENTARY MATERIALS L. G. ZAVILEYSKIY, ET AL. "CLASSIFICATION AND QUANTIFICATION OF UNPRODUCTIVE SPLICING EVENTS"

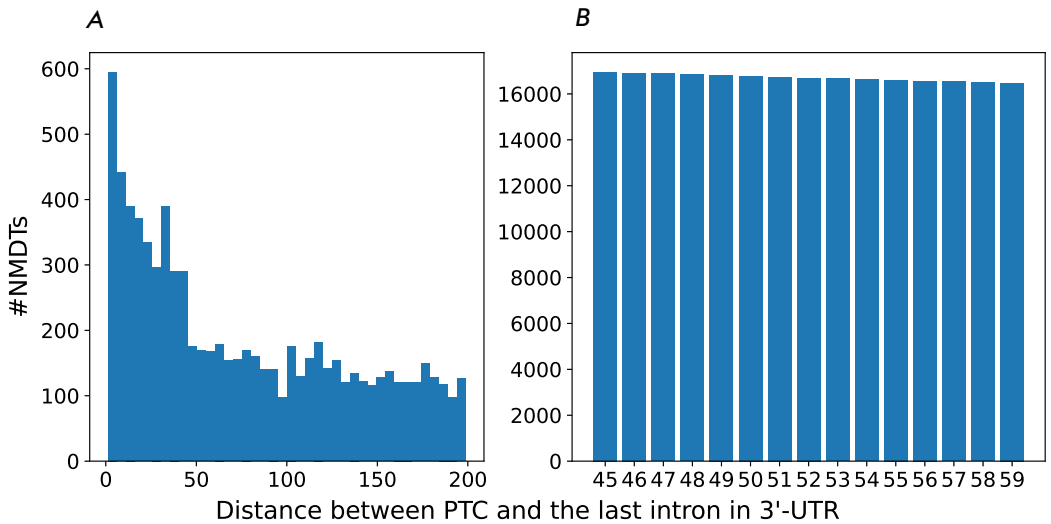


Fig. S1. Histogram of the distances between the stop codon and the last intron in the 3′-UTR of annotated human transcripts (*A*). The number of predicted NMDTs as a function of the chosen threshold for the distance from the stop codon to the intron (*B*)

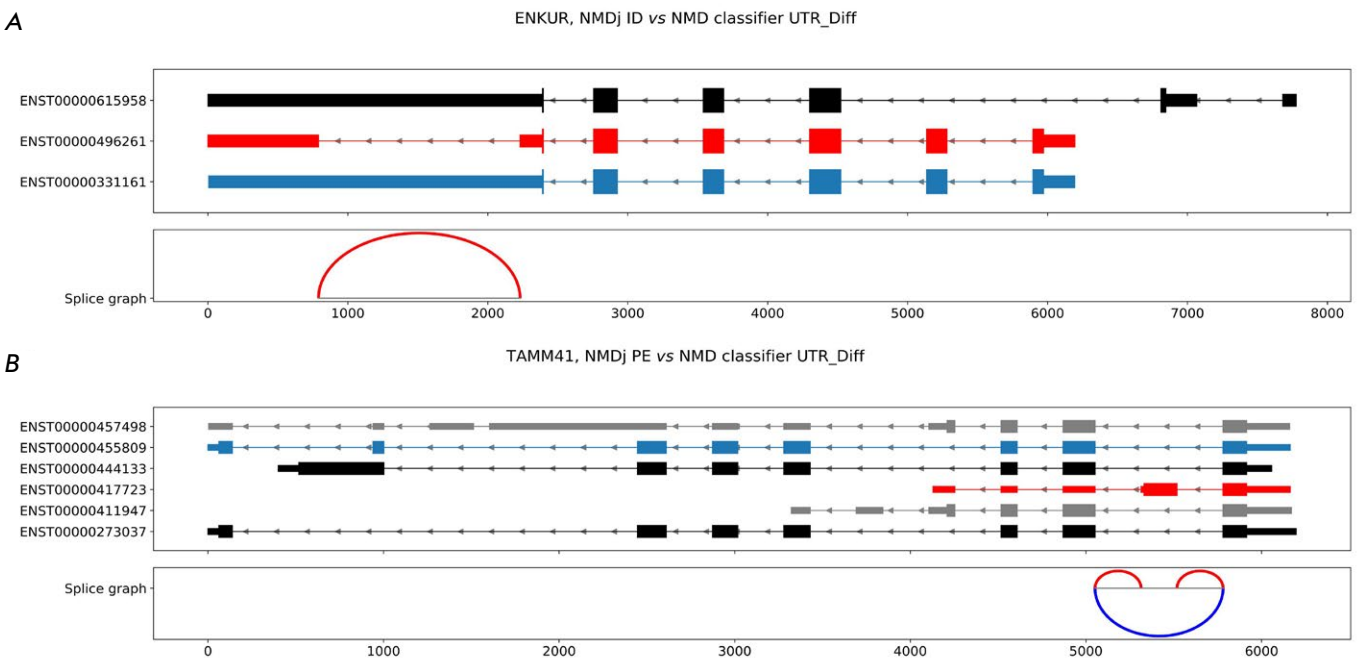


Fig. S2. Examples of a detailed classification of AS events categorized as UTR_Diff by NMD Classifier: intron detention (*A*) and poison exon (*B*). MANE-Select transcripts are shown in blue, NMDTs of interest are in red, other NMDTs and protein-coding transcripts are in gray and black, respectively

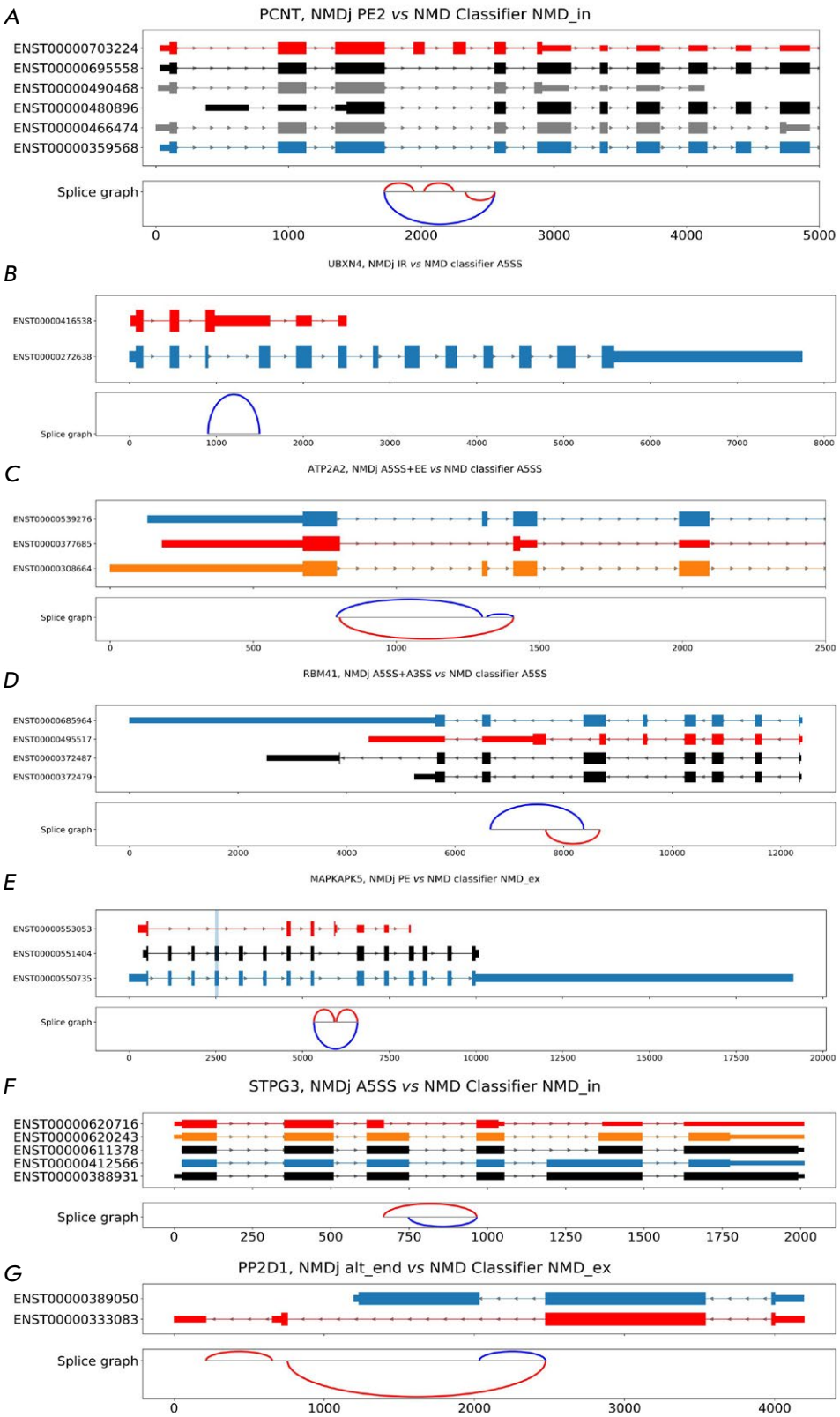


Fig. S3. Examples of misclassification by NMD Classifier. MANE-Select transcripts are shown in blue, NMDTs of interest are in red, other NMDTs and protein-coding transcripts are in gray and black, respectively. Best partner transcripts are shown in orange in the case where they are not similar to the MANE-Select transcript