БИОХИМИЯ, 2023, том 88, вып. 10, с. 2014 – 2031

УДК 577.122

ПРОТЕОМНЫЙ ПОДХОД ДЛЯ ИССЛЕДОВАНИЯ ЭКСПРЕССИИ, ЛОКАЛИЗАЦИИ И ФУНКЦИЙ БЕЛКОВОГО ПРОДУКТА ГЕНА *SOWAHD* В ПРОЦЕССЕ ГРАНУЛОЦИТАРНОЙ ДИФФЕРЕНЦИРОВКИ

© 2023 С.Е. Новикова*, Т.В. Толстова, Н.А. Соловьева*, Т.Е. Фарафонова, О.В. Тихонова, Л.К. Курбатов, А.Л. Русанов, В.Г. Згода*

НИИ биомедицинской химии имени В.Н. Ореховича, 119121 Москва, Россия; электронная почта: novikova.s.e3101@gmail.com, victor.zgoda@gmail.com

Поступила в редакцию 22.06.2023
После доработки 23.08.2023
Принята к публикации 26.08.2023

Каталогизация белков человека, определение уровня их содержания, клеточной локализации, выполняемой функции и потенциальной медицинской значимости являются важными задачами, стоящими перед мировым протеомным сообществом. В настоящее время локализация и функции белковых продуктов для почти половины белок-кодирующих генов неизвестны или слабо изучены. Исследование протеома органелл является многообещающим подходом для выявления локализации и функций белков человека. Протеом клеточного ядра представляет особый интерес, поскольку многие белки с ядерной локализацией, например, транскрипционные факторы, выполняют регуляторные функции, определяющие судьбу клеток. По результатам метаанализа ядерного протеома, или нуклеома, клеток линии HL-60 под действием полностью-*транс*-ретиноевой кислоты (ATRA) выявили, что функция и локализация белкового продукта гена *SOWAHD* слабо изучена, кроме того, отсутствует исчерпывающая информация об экспрессии *SOWAHD* на уровне белка. В клетках линии HL-60 для белок-кодирующего гена *SOWAHD* определили экспрессию мРНК на уровне 6,4 ± 0,7 транскриптов на миллион молекул. С помощью направленной масс-спектрометрии измерили содержание белка SOWAHD (белок 58, содержащий домен анкириновых повторов) в диапазоне 0,27–1,25 фмоль/мкг общего белка. С применением пульсового мечения стабильными изотопами определили, что время полураспада для белкового продукта гена *SOWAHD* составляет приблизительно 19 ч. Протеомное профилирование ядерной фракции клеток линии HL-60 показало, что содержание белка SOWAHD увеличивалось в процессе ATRA-индуцированной гранулоцитарной дифференцировки, с пиком в 9 ч после добавления индуктора и с последующим снижением в более поздние временные точки. Результаты исследования впервые указывают на ядерную локализацию и вовлечённость белкового продукта гена *SOWAHD* в индуцированную гранулоцитарную дифференцировку.

КЛЮЧЕВЫЕ СЛОВА: белковый продукт гена *SOWAHD*, транскриптом, протеом, ATRA, ядро клетки, масс-спектрометрия, пульсовое мечение стабильными изотопами, динамика синтеза и деградации белка.

DOI: 10.31857/S0320972523100196, **EDN:** OYBYYU

ВВЕДЕНИЕ

Белки представляют собой конечные продукты экспрессии генов. Именно они выполняют различные функции в живых клетках, опосредуя их жизнедеятельность. Каталогизация, оценка уровня экспрессии, определение клеточной локализации и выполняемой функции, а также выявление потенциальной биологической и медицинской значимости белков являются основными задачами международного проекта «Протеом человека», запущенного в 2010 г. В рамках проекта для регистрации экспрессии белков и определения её уровня,

Принятые сокращения: клетки линии HL-60 – клетки острого миелоидного лейкоза; ATRA – полностью-*транс*-ретиноевая кислота; SILAC – стабильное изотопное мечение клеток аминокислотами в культуре; SIS – меченные стабильными изотопами пептидные стандарты; SOWAHD – белок 58, содержащий домен анкириновых повторов; SRM – мониторинг выбранных реакций; TMT – изобарные тандемные метки.

* Адресат для корреспонденции.

т.е. для каталогизации, применяют масс-спектрометрию и используют моноклональные антитела, а одним из практических результатов является создание баз знаний, таких как Uniprot [1], NextProt [2], Human Protein Atlas [3] и т.д., содержащих исчерпывающую информацию о протеоме человека [4, 5].

По разным источникам, у человека на сегодняшний день насчитывается от 19 379 до 21 306 белок-кодирующих генов [6]. С точки зрения экспрессии, согласно критериям международной организации «Протеом человека» (Human Proteome Organization (HUPO)), продукты потенциальных белок-кодирующих генов в зависимости от доказанности их существования на уровне протеома (protein existence, PE) можно разделить на 5 групп (PE1–PE5) [5]. Наиболее надёжно подтверждённые белковые продукты получают статус PE1, в то время как остальные (PE 2, 3, 4, 5) представляют собой так называемые «недостающие/упущенные» (missing) белки. Для их белок-кодирующих генов ранее могли быть обнаружены транскрипты у человека (PE2) или могли быть предсказаны ортологи белков, подтверждённых у других биологических видов (PE3), или белковый продукт был предсказан с помощью биоинформатических моделей (PE4), при этом подтверждённые данные об экспрессии белка у человека отсутствуют. Согласно промежуточным итогам проекта «Протеом человека», подведённым в 2020 г., именно за счёт обнаружения «недостающих/упущенных» белков покрытие протеома увеличилось с 69,8 до 90,4% от числа всех белок-кодирующих генов [5].

Масс-спектрометрический анализ благодаря высокой производительности и возможности своего рода «секвенирования» участков аминокислотной последовательности хорошо зарекомендовал себя для обнаружения «недостающих/упущенных» белков [2, 7, 8]. Успешное детектирование таких белков было, с одной стороны, продиктовано развитием масс-спектрометрического метода, в том числе совершенствованием масс-анализаторов, применением улучшенных способов пробоподготовки, использованием фракционирования и изобарных меток. С другой стороны, привлечение специфических типов биологического материала, таких как плаценты, репродуктивные и эмбриональные ткани, позволило зарегистрировать «недостающие/упущенные» белки [9]. Исследование белкового состава отдельных органелл в качестве потенциального источника «недостающих/упущенных» белков также заслуживает внимания как один из подходов для углубления знания о протеоме человека. Особый интерес представляет протеомный состав клеточного ядра.

Подтверждение экспрессии белковых продуктов или каталогизация, выполненная на данный момент времени для более чем 90% белок-кодирующих генов [5], является лишь одной из задач, поставленных перед проектом «Протеом человека». Более сложным

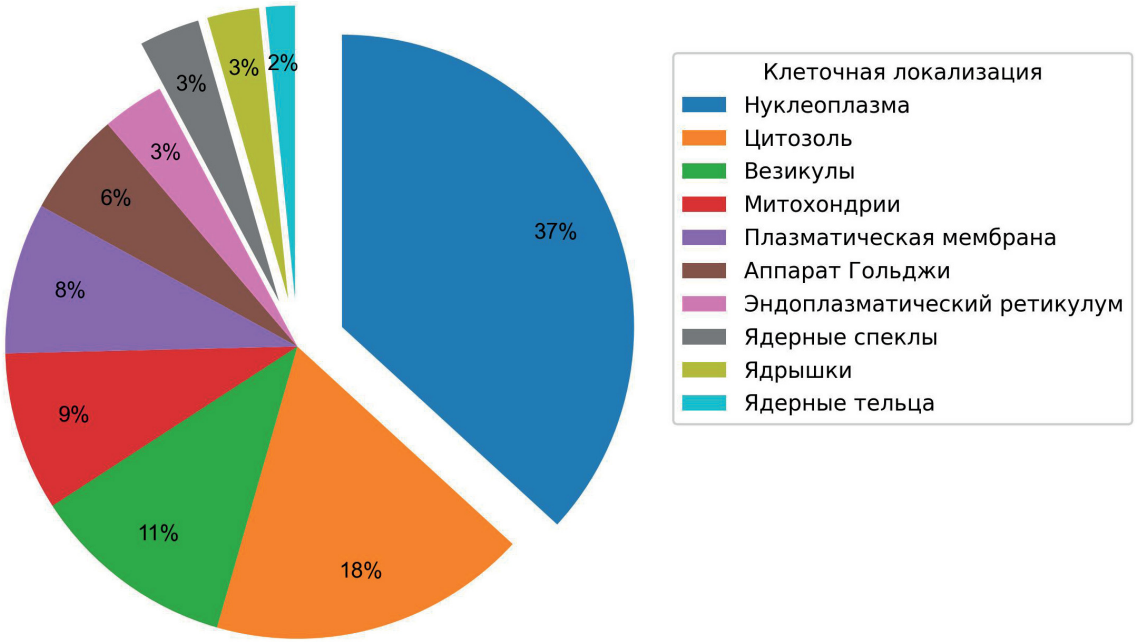

Рис. 1. Круговая диаграмма, демонстрирующая распределение белков с известной клеточной локализацией согласно данным Human Proteome Atlas (версия 22.0). Показаны первые десять групп с наибольшим количеством белков. Стоит отметить, что среди белков с известной клеточной локализацией примерно треть обнаруживается в ядре

представляется определение локализации, функции и медицинской значимости белков с уже установленной экспрессией. Для многих белков такая информация отсутствует в основных протеомных базах данных.

Клеточную локализацию белков можно определить с помощью визуализации их комплексов с флюоресцентно мечеными моноклональными антителами методом микроскопии. Основные результаты, полученные с помощью такого подхода, обобщены в базе данных Human Proteome Atlas [10]. Версия этого ресурса, выпущенная в 2022 г. (версия 22.0), содержит данные для 17 200 белок-кодирующих генов человека, что составляет примерно 86% от их общего числа. При этом только для 13 105 (около 61%) известна информация о клеточной локализации (рис. 1).

Определение функции белка является наиболее сложной задачей. Зачастую роль белка изучают в контексте определённых заболеваний, чаще всего наиболее социально значимых, таких как онкологические и кардиоваскулярные заболевания, а также болезни центральной нервной системы. В ходе выполнения проекта «Протеом человека» активно используется так называемый хромосомо-центричный подход, когда стоит задача в первую очередь доказать экспрессию всех белков, кодируемых каждой хромосомой. В таком случае накопление информации об экспрессии белкового продукта может значительно опережать поступление данных о функциональных особенностях белков и их медицинской значимости.

Белки, входящие в состав ядра, представляют особый интерес, поскольку к ним относятся важные регуляторы – транскрипционные факторы, коактиваторы и корепрессоры транскрипции, ферменты, изменяющие эпигенетический статус клетки, белки, принимающие участие в репликации и репарации ДНК [11]. Исследование протеома ядра в модели индуцированной гранулоцитарной дифференцировки позволило выявить динамическую пертурбацию регуляторных молекул, в том числе и транскрипционных факторов [12, 13].

Мы провели поиск «недостающих/упущенных» белков, а также белков, функция и локализация которых плохо исследованы, используя масс-спектрометрические данные для ядер клеток линии острого миелоидного лейкоза (HL-60) под действием полностью-*транс*-ретиноевой кислоты (ATRA), опубликованные нами ранее [13]. Для белкового продукта гена *SOWAHD* информация об экспрессии в основных базах данных оказалась противоречивой. В базе данных Uniprot указана информация об экспрессии продукта гена *SOWAHD* только на уровне транскрипта (PE2), в то же время, по материалам базы данных NextProt, белок SOWAHD (белок 58, содержащий домен анкириновых повторов) был зарегистрирован масс-спектрометрическим методом в ряде протеомных экспериментов (PE1), в основном в клетках крови, ткани плаценты и тестикулах. При этом, согласно основным протеомным базам данных, включая наиболее детальную базу знаний NextProt, локализация, функция и медицинская значимость белка SOWAHD неизвестны. В результате для белкового продукта гена *SOWAHD* осуществили всесторонний транскриптомный и протеомный анализ для разъяснения его локализации и вероятной биологической функции, а также времени полураспада белкового продукта в процессе гранулоцитарной дифференцировки.

МАТЕРИАЛЫ И МЕТОДЫ

Для количественной оценки белка SOWAHD были использованы данные протеомного профилирования с использованием изобарных тандемных меток (TMT, Tandem Mass Tag), опубликованные ранее [13]. Эксперименты по направленной масс-спектрометрии, пульс-мечению SILAC, транскриптомному анализу проведены непосредственно для данного исследования.

Метаанализ данных по протеомике ядра. Для белков, ранее идентифицированных масс-спектрометрическим методом в ядре клеток линии HL-60 под действием ATRA, получили информацию об аннотационном статусе, биологической функции, клеточной локализации, основных публикациях и об аннотации по категориям базы знаний GeneOntology (GO) из базы данных Uniprot (версия 2023_01).

Культивирование клеток линий HL-60, NB4, K562, HepG2, HEK293 и HeLa. Культуры клеток линий HL-60 и K562 были получены из криобанка «Коллекция культур клеток позвоночных» Института цитологии РАН (Санкт-Петербург, Россия). Линия NB4 была закуплена в компании «CLS Cell Lines Service GmbH» (Германия). Культуры клеток линий HepG2, HEK293 и HeLa были взяты из криобанка НИИ биомедицинской химии им. В.Н. Ореховича (ИБМХ, Москва, Россия). Идентичность всех клеточных линий была подтверждена ДНК-профилем (анализ коротких тандемных повторов, STR).

После разморозки клетки 3 раза отмывали от DMSO в ростовой среде RPMI-1640

(«Gibco», США) с добавлением фетальной сыворотки крупного рогатого скота (ФБС) до конечной концентрации 10% (*v/v*) с помощью центрифугирования. Культуры культивировали в ростовой среде RPMI-1640 с добавлением ФБС до конечной концентрации 10% (*v/v*) в присутствии 100 ед./мл пенициллина, 100 ед./мл стрептомицина и 2 мМ L-глутамина (все реактивы – «Gibco») в CO_2-инкубаторе в стандартных условиях (37 °C, 5% CO_2, 80% влажности) в течение 7 суток. При достижении концентрации 1 млн кл./мл культуры рассеивали в соотношении 1 : 3. Подсчёт клеток осуществляли в камере Горяева.

Индуцированная гранулоцитарная дифференцировка клеток линии HL-60. Для дифференцировки клетки линии HL-60 помещали в 15 мл ростовой среды в концентрации 0,3 млн кл./мл и добавляли ATRA («Sigma-Aldrich», США) до конечной концентрации 10 мкМ.

Изучение динамики протеома в процессе индуцированной гранулоцитарной дифференцировки ранее проводилось по различным временным схемам: 0, 12 и 48 ч [14]; 0, 24, 48, 72 и 96 ч [15]; 0, 3, 24 и 96 ч [12]; 0, 3, 24, 48 и 96 ч [16]. Поверхностный маркер гранулоцитарной дифференцировки CD11b в разы увеличивает свою экспрессию уже к 24 ч после обработки ATRA [17], а к 72 ч клетки приобретают фенотип зрелых гранулоцитов (нейтрофилов). При этом изменения в протеоме, которые в том числе ответственны за экспрессию специфических белков нейтрофилов, могут происходить на начальных стадиях гранулоцитарной дифференцировки. Чтобы выявить ранние пертурбации протеома, мы взяли временные точки 0, 3, 6, 9 и 12 ч, а также точку 72 ч как ту, в которой клетки приобретают фенотип зрелых гранулоцитов под воздействием индуктора.

Для протеомного анализа ядерных фракций клетки осаждали с помощью центрифугирования и 2 раза промывали осадок фосфатно-солевым буфером (PBS) («Sigma-Aldrich»). В качестве контроля использовали интактные клетки линии HL-60. Для протеомного эксперимента клетки линии HL-60 были индуцированы ATRA в трёх биологических повторах. Эксперименты по дифференцировке были проведены в пределах 15–20 пассажей от вывода клеток из заморозки.

Культивирование интактных клеток линии HL-60 для введения изотопно-меченных аминокислот лизина и аргинина. Динамику синтеза и деградации белка SOWAHD определяли с помощью пульсового мечения стабильными изотопами [18]. В основе лежит технология стабильного изотопного мечения клеток аминокислотами в культуре (SILAC, Stable Isotope Labeling by/with Amino acids in Cell culture). При использовании данного подхода клетки культивируют в присутствии так называемых «тяжёлых» аминокислот, содержащих атомы стабильных изотопов ^{13}C и ^{15}N. В клетках «тяжёлые» аминокислоты используются при синтезе белка вместо нативных, или «лёгких». Масс-спектрометрический анализ позволяет оценить соотношение «лёгких» и «тяжёлых» пептидов в различные моменты времени культивирования. Данное соотношение используют для построения математической модели, описывающей кинетику деградации белка. Такой моделью является кинетическое уравнение реакции первого порядка [19, 20]. В данном исследовании пульсовое мечение SILAC комбинировали с направленным масс-спектрометрическим анализом в режиме мониторинга выбранных реакций (Selected Reaction Monitoring, SRM) вместо панорамной масс-спектрометрии. Такой подход позволяет проводить анализ низкокопийных белков, а также обеспечивает более точное определение времени полураспада белка [21]. Время полураспада белка SOWAHD ($t_{1/2}$) вычисляли по уравнению аппроксимирующей функции для зависимости деградации белка SOWAHD от времени с учётом времени удвоения клеток линии HL-60 (48 ч).

На основании аппроксимирующей функции зависимости деградации белка SOWAHD от времени определяли общее время полураспада, не учитывающее деление клеток, по следующей формуле:

$$t_{1/2общ} = e^{(L0-b)/a},$$

где L0 – ½ от значения интенсивности дочернего иона (MS2, фрагмент y10+, −GGPSGP GSSR) пептида AEELGGPSGPGSSR во временной точке 0 ч, a и b – коэффициенты аппроксимирующей функции.

Время полураспада белка корректировали с учётом деления клеток: вычисляли общую константу деградации ($k_{общ}$) по формуле $k_{общ} = \ln 2/t_{1/2общ}$ и константу деления с учётом времени удвоения ($t_{удв}$) клеток линии HL-60 (48 ч) по формуле $k_{дел} = \ln 2/t_{удв}$, вычисляли время полураспада белка ($t_{1/2}$) по формуле $t_{1/2} = \ln 2/(k_{общ} - k_{дел})$.

Время полураспада белкового продукта SOWAHD ($t_{1/2}$) было рассчитано отдельно для каждого их трёх технических повторов с использованием аппроксимирующей функции для зависимости деградации белка SOWAHD

от времени, с последующим вычислением среднего значения. Результаты приведены в таблице 1П Приложения 1 (лист «SRM SILAC»).

После выхода из заморозки клетки линии HL-60 культивировались в условиях, описанных выше. Затем интактные клетки линии HL-60 были помещены в среду культивирования из набора SILAC Protein Quantitation Kit (Trypsin) RPMI 1640 («Thermo Fisher Scientific», США), в которую были добавлены «тяжёлые» аминокислоты 13C6 15N2 L-лизин-2 HCl и 13C6 15N4 L-аргинин HCl («Thermo Fisher Scientific») в количестве, рекомендованном производителем, а также диализованная ФБС («Gibco») до конечной концентрации 10% (*v/v*). Начальная концентрация клеток в среде составляла 0,15 млн кл./мл. Далее клетки культивировали в CO_2-инкубаторе в стандартных условиях (37 °C, 5% CO_2, 80% влажности). Для протеомного анализа клетки отбирали через 0, 1, 6, 9, 24, 72 и 120 ч, после чего осаждали с помощью центрифугирования и 2 раза промывали PBS («Sigma-Aldrich», США).

Выделение общей РНК и полногеномное транскриптомное секвенирование линий HL-60, NB4, K562, HepG2, HEK293 и HeLa. Общую РНК выделяли при помощи фирменного набора RNeasy Mini Kit («QIAGEN», США) согласно рекомендациям производителя. Качество выделенной РНК определяли на приборе Agilent 2100 Bioanalyser («Agilent Technologies», США) с использованием чипов RNA 6000 Nano LabChip («Agilent Technologies»). Концентрацию общей РНК определяли спектрофотометрически по поглощению при длине волны 260 нм на микроспектрофотометре ND-1000 («Thermo Fisher Scientific»). Секвенирование было выполнено на платформе NovaSeq 6000 Illumina с длиной прочтений 100 пар оснований в трёх технических повторах. Исходные данные РНК-секвенирования в процессе обработки претерпевают нормализацию на основании длины гена и глубины секвенирования. В результате экспрессия генов выражена в количестве транскриптов на миллион (TPM). Сумма TPM экспрессии для всех генов оказывается одинаковой для различных образцов, что предоставляет возможность сравнивать между собой экспрессионные данные [22]. Данные экспрессии генов семейства SOWAH приведены в таблице 1П Приложения 1 (лист «Транскриптомные данные»).

Пробоподготовка клеток линии HL-60 к протеомному профилированию ядерных фракций. Для выделения ядерной фракции мы использовали протокол с участием градиента сахарозы, применённый нами ранее [23]. К осадку клеток (приблизительно 5 млн клеток) добавляли 200 мкл буфера, содержащего 10 мМ HEPES-NaOH (pH 7,9), 1,5 мМ $MgCl_2$, 10 мМ KCl, 0,1 мМ ЭДТА, 0,32 М сахарозу, а также коктейль ингибиторов протеаз (все реагенты – «Sigma Aldrich»), и 3–5 раз перемешивали пипетированием. Затем добавляли ещё 200 мкл вышеописанного буфера, содержащего, помимо прочих компонентов, 0,5%-ный (*m/v*) NP-40 («Pierce», США), 3–5 раз перемешивали и инкубировали 10 мин во льду. Полученную суспензию наслаивали на 800 мкл буфера, содержащего 10 мМ HEPES-NaOH (pH 7,9), 1,5 мМ $MgCl_2$, 10 мМ KCl, 0,1 мМ ЭДТА, 1,2 М сахарозу, а также коктейль ингибиторов протеаз, и затем центрифугировали в течение 20 мин со скоростью 16 000 *g* при температуре 4 °C. К осадку, представляющему собой ядерную фракцию, добавляли 100 мкл лизирующего буфера, содержащего 1%-ный (*m/v*) SDS и 100 мМ Tris-HCl (pH 8,5) (все реагенты – «Sigma-Aldrich»), после чего подвергали воздействию ультразвука с помощью дезинтегратора со щупом («Bandelin Sonopuls», Германия) с мощностью 50% в течение 2 мин во льду. Затем смесь центрифугировали в течение 10 мин со скоростью 16 000 *g* при температуре 4 °C. Супернатант, содержащий белки ядерной фракции, отбирали и сохраняли для дальнейшего анализа. В полученных образцах определяли концентрацию общего белка колориметрическим методом с помощью коммерческого набора Pierce™ BCA Protein Assay Kit («Pierce») в соответствии с рекомендациями производителя. При выделении ядерной фракции с помощью центрифугирования целевые белки могут быть контаминированы белками других органелл, например, митохондрий. Аннотация нуклеома клеток линии HL-60, идентифицированного в ходе гранулоцитарной дифференцировки по категориям «Внутриклеточная локализация» базы данных «GeneOntology», показала контаминацию белками митохондрий на уровне 11,5%.

Гидролитическое расщепление белков. Гидролитическое расщепление белков осуществляли согласно протоколу FASP (Filter-Aided Sample Preparation) [24] с некоторыми изменениями. Для восстановления и алкилирования дисульфидных связей цистеина к образцам лизата клеток, содержащим 100 мкг общего белка, добавляли трис(2-карбоксиэтил)фосфин(TCEP) и 2-хлорацетамид (CAA) («Sigma-Aldrich») до конечных концентраций 20 мМ и 32 мМ соответственно. Затем пробы инкубировали при температуре 80 °C в течение 40 мин, после чего образцы наносили на кон-

центрирующие фильтры с отсечением 30 кДа и 3 раза промывали буфером, содержащим 8 М мочевины в 0,1 М Tris-HCl (pH 8,5), путём центрифугирования при скорости 11 000 *g* в течение 15 мин при температуре 20 °С. Затем образцы 3 раза промывали буфером, содержащим 0,1 М Tris-HCl (pH 8,5), и 2 раза – буфером для трипсинолиза (50 мМ ТЭАБ, pH 8,5), путём центрифугирования при скорости 11 000 *g* в течение 15 мин и температуре 20 °С. Затем к каждому образцу добавляли 50 мкл 50 мМ ТЭАБ (pH 8,5), а также раствор трипсина («Promega», США) с концентрацией 0,5 мкг/мкл в соотношении массы трипсина к массе общего белка 1 : 50, после чего пробу инкубировали в течение ночи при температуре 37 °С. По окончании инкубации пептиды элюировали путём центрифугирования при скорости 11 000 *g* в течение 15 мин при температуре 20 °С и промывали фильтр 100 мкл 5%-ной (*v/v*) муравьиной кислоты. Общее содержание пептидов определяли с помощью набора Pierce Quantitative Peptide Assays («Pierce») в соответствии с рекомендациями производителя. Половину от общего количества пептидов замораживали и хранили на −80 °С, затем данную аликвоту использовали для SRM-измерений, проведённых в данной работе. Другая половина была использована в экспериментах с TMT-метками, описанных ранее [13].

Панорамный и направленный масс-спектрометрический анализ в режиме SRM с использованием меченных стабильными изотопами пептидных стандартов (SIS). Панорамный масс-спектрометрический анализ был выполнен, как описано ранее [13]. Масс-спектрометрические данные анализировали с помощью программного обеспечения MaxQuant 2.0.3.0 со встроенным поисковым алгоритмом Andromeda. Идентификацию проводили относительно FASTA-файла, содержащего аминокислотные последовательности белков человека (2022/02) и его инвертированного аналога для вычисления частоты ложноположительных идентификаций (FDR). Для белков и пептидов пороговое значение FDR = 0,01.

Количественный анализ осуществляли на основании интенсивности репортерных ионов, образующихся при фрагментации изобарных TMT-меток, с помощью встроенного в MaxQuant алгоритма. Применяли нормализацию по взвешенной медиане с использованием каналов для всех TMT_i (i = 10) репортерных ионов в качестве референсных. Применяли функцию сопоставления пептидных ионов между ЖХ-МС/МС-экспериментами (Matching between runs, MBR). Дополнительно осуществляли нормализацию по суммарному сигналу всех белков в каждом TMT_i-канале. Данные панорамного анализа доступны через репозиторий PRIDE (идентификатор PXD035281) и репозиторий MassIVE (идентификатор MSV000092670).

С помощью направленной масс-спектрометрии в режиме SRM с использованием SIS-пептидов (SRM/SIS) было измерено содержание протеотипических пептидов AAPAGWLSEER и AEELGGPSGPGSSR, картируемых на белок SOWAHD. Критериями выбора пептидов для синтеза стандартов были: длина 9–20 аминокислотных остатков; отсутствие в последовательности цистеина (C), метионина (M), *N*-концевых глутаминовой кислоты (E) и глутамина (Q), а также пропущенных сайтов гидролиза; уникальность была проверена с помощью ресурса Peptide unicity checker – neXtProt. Выбранные для таргетного анализа пептиды также представлены в учётной записи Nextprot (в разделе https://www.nextprot.org/entry/NX_A6NJG2/peptides). Твердофазный пептидный синтез SIS-стандартов осуществляли на автоматическом пептидном синтезаторе Overture («Protein Technologies», Великобритания). При синтезе изотопно-меченных пептидов вместо обычного лизина или аргинина использовали изотопно-меченные аминокислоты Fmoc-Lys-OH-$^{13}C_6$,$^{15}N_2$ или Fmoc-Arg-OH-$^{13}C_6$,$^{15}N_2$ («Cambridge Isotope Laboratories», США).

Направленный масс-спектрометрический анализ в режиме SRM проводили по следующему протоколу. Каждый экспериментальный образец анализировали в трёх технических повторах. Анализ SRM проводили для тех же образцов, которые использовались в панорамном масс-спектрометрическом эксперименте. Хроматографическое разделение осуществлялось с помощью системы Agilent 1200 series («Agilent Technologies»), соединённой с тройным квадрупольным масс-анализатором TSQ Quantiva («Thermo Fisher Scientific»). Образец объёмом 3,5 мкл, содержащий 15 мкг нативных пептидов и стандарты SIS, разделяли с использованием аналитической колонки ZORBAX SB-C18 (150 × 0,5 мм, диаметр частиц 5 мкм, «Agilent Technologies») в градиенте ацетонитрила со скоростью потока 20 мкл/мин. В начале колонку уравновешивали 5%-ным (*v/v*) раствором B (80%-ный (*v/v*) ацетонитрил в 0,1%-ной (*v/v*) муравьиной кислоте) и 95%-ным (*v/v*) раствором A (0,1%-ная (*v/v*) муравьиная кислота) в течение 5 мин, затем концентрацию раствора B линейно увеличивали до 60% (*v/v*) за 30 мин, после чего

концентрацию раствора B увеличивали до 99% (*v*/*v*) за 1 мин и в течение 5 мин промывали колонку 99%-ным (*v*/*v*) раствором B, затем в течение 1 мин концентрация возвращалась к начальным условиям, в которых колонка уравновешивалась в течение 9 мин. Масс-спектрометрический анализ осуществляли в режиме динамического мониторинга выбранных реакций (dSRM), используя следующие настройки МС-детектора: напряжение на капилляре составляло 4000 В, скорость осушающего газа (азот) составляла 7 л/мин, скорость аксилярного газа (азот) составляла 5 л/мин, температура капилляра – 350 °C, окно изоляции для первого и третьего квадруполя – 0,7 Да, время цикла сканирования – 1,2 с, давление газа (аргон) в ячейке соударения – 1,5 мТорр. Окно времени удерживания на колонке обращённой фазы составило 3 мин для каждого прекурсорного иона. Для количественной оценки в ПО Skyline (версия 4.1.0) использовались все транзиции, сигнал от которых регистрировался. Для каждой пробы содержание белка определялось как содержание соответствующего ему уникального протеотипического пептида, измеренного в трёх биологических повторах. Калибровочные кривые для стандартных пептидов (STD/SIS) AEELGGPSGPGSSR и AAPAGWLSEER, картируемых на белок SOWAHD, приведены на рис. 1П Приложения 2.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ

Метаанализ нуклеома клеток линии HL-60. На рис. 2 показаны результаты метаанализа для белков, ранее идентифицированных масс-спектрометрическим методом в ядре клеток линии HL-60 (1860 белков) (нуклеом). Используя базу знаний Uniprot (версия 2023_01), протеомные данные фильтровали на основании аннотационного балла Uniprot < 5 (117 белков), затем на каждой последующей стадии оставляли белки с неизвестной биологической функцией (53 белка), с неизвестной клеточной локализацией (26 белков), белки, для которых в Uniprot не представлено релевантных публикаций, и, наконец, белки, для которых отсутствовала аннотация GO (2 белка).

По результатам фильтрации удалось выявить 2 белка, удовлетворяющих всем условиям: ANKRD58 (ankyrin repeat domain 58), кодируемый геном *SOWAHD*, и неохарактери-

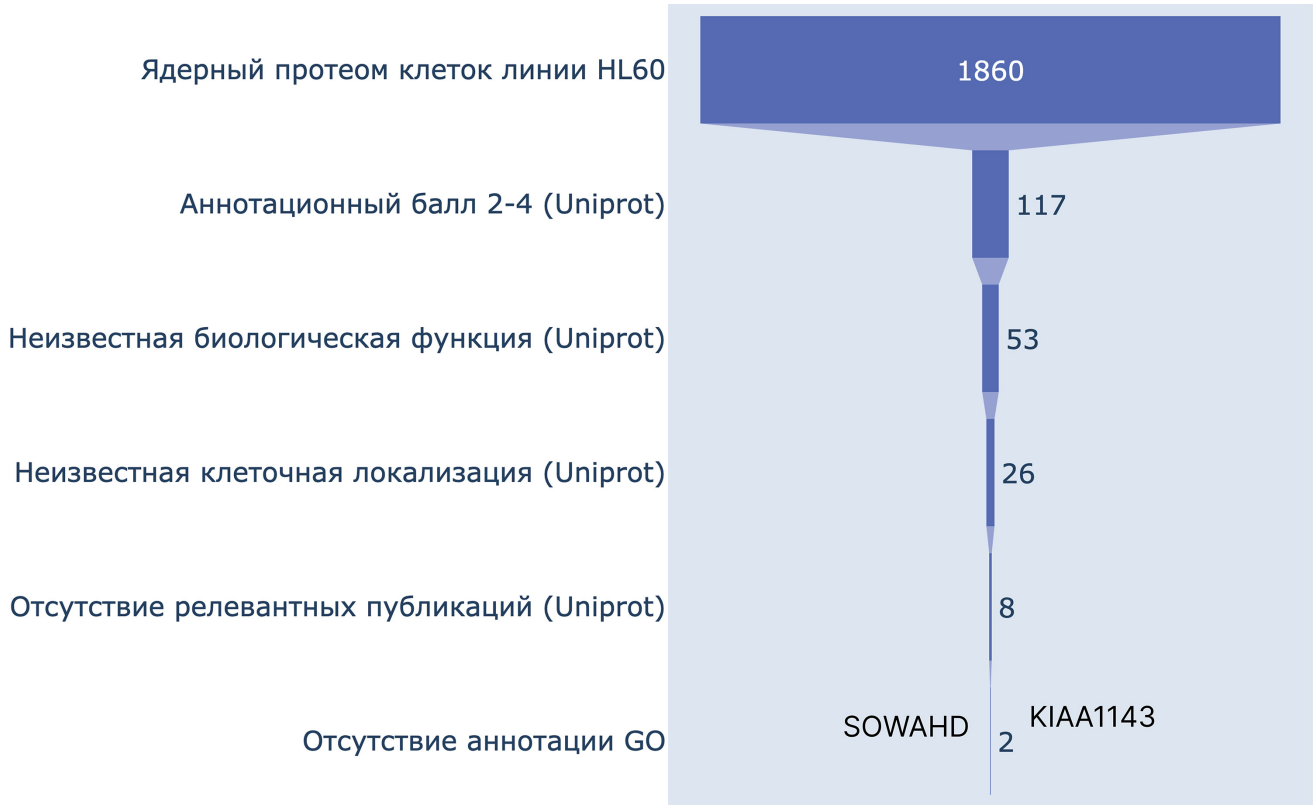

Рис. 2. Схема фильтрации масс-спектрометрических данных, полученных для ядерного протеома клеток линии HL-60 (1860 белков), на основании аннотационного статуса, биологической функции, клеточной локализации, информации об основных публикациях и аннотации по категориям базы знаний GO

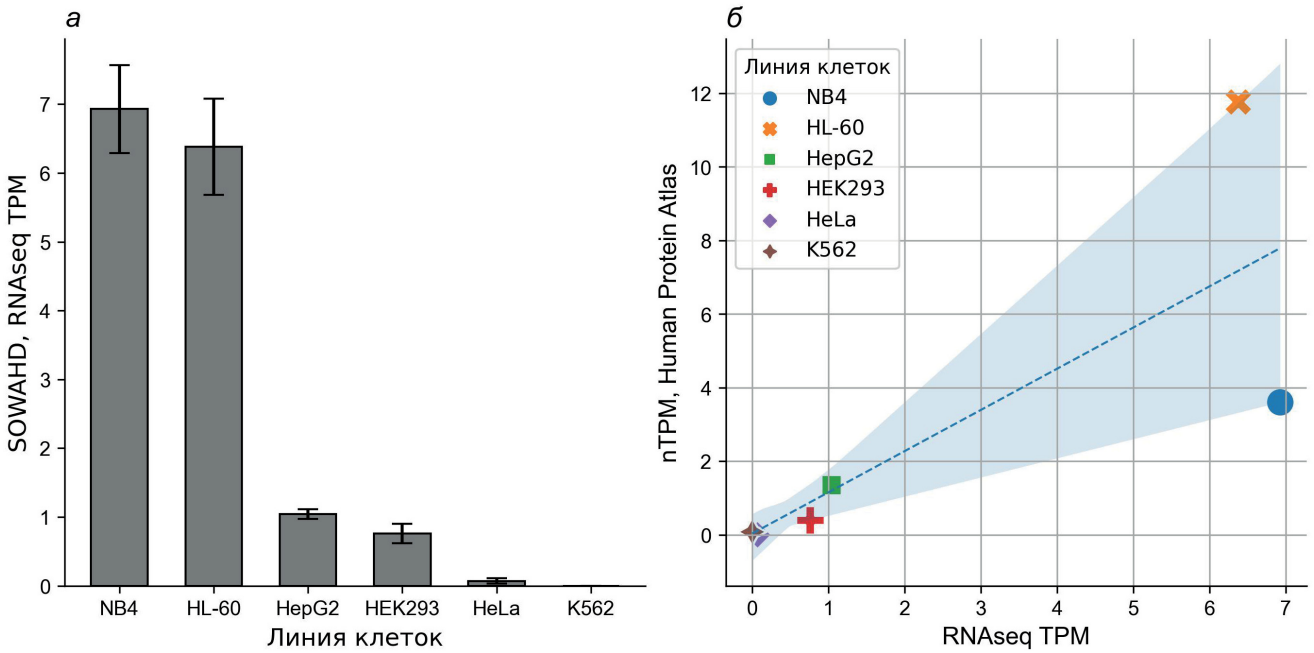

Рис. 3. Результаты транскриптомного анализа. *а* – Уровень экспрессии мРНК SOWAHD, измеренный с помощью полногеномного РНК-секвенирования (RNAseq) в клетках линий NB4, HL-60, HepG2, HEK293, HeLa и K562. По оси *y* отложено количество транскриптов на миллион, TPM (transcripts per million). *б* – Корреляция наших экспериментальных данных и материалов базы знаний Protein Atlas (версия 22.0) для уровня экспрессии мРНК SOWAHD в клетках линий NB4, HL-60, HepG2, HEK293, HeLa и K562. По оси *y* отложено нормализованное количество транскриптов на миллион, nTPM, материалы ProteinAtlas (версия 22.0); по оси *x* – количество TPM, определённое в нашем эксперименте

зованный белок KIAA1143 (ген *KIAA1143*, код Uniprot Q96AT1). Для белка SOWAHD в базе данных Uniprot экспрессия указана только на уровне мРНК (PE2). В то же время для гена *KIAA1143* экспрессия на белковом уровне подтверждена во всех основных протеомных базах данных (PE1). В структуре белка SOWAHD имеются анкириновые повторы, за счёт которых осуществляются белок-белковые взаимодействия, что может указывать на регуляторную роль SOWAHD. Дальнейшие исследования проводили для данного белка.

Экспрессия транскрипта SOWAHD в линиях клеток гемопоэтического и эпителиального происхождения. Для оценки экспрессии мРНК SOWAHD мы использовали экспериментальные данные полногеномного РНК-секвенирования (RNAseq), полученные для гемопоэтических клеток линий HL-60, NB4 и K562, а также линий клеток эпителиального происхождения HeLa, HepG2 и HEK293. Кроме того, сравнивали результаты нашего транскриптомного эксперимента с данными базы знаний ProteinAtlas (версия 22.0). Результаты показаны на рис. 3.

Как видно из рис. 3, *а*, наибольшую экспрессию мРНК SOWAHD на уровне 6,9 ± 0,6 и 6,4 ± 0,7 TPM наблюдали в клетках острого миелоидного лейкоза линий NB4 и HL-60 соответственно. В клетках хронического миелогенного лейкоза линии K562 экспрессия мРНК SOWAHD оказалась ниже порога детекции. Умеренная экспрессия мРНК SOWAHD на уровне 1,1 ± 0,1 и 0,8 ± 0,1 TPM была зарегистрирована в клетках эпителиального происхождения линий HepG2 (ткань печени) и HEK293 (эмбриональные почки). Экспрессию мРНК SOWAHD на уровне 0,07 ± 0,03 TPM наблюдали в клетках эпителиального происхождения линии HeLa (ткань шейки матки).

Рисунок 3, *б* показывает корреляцию (R^2 = = 0,63) наших экспериментальных результатов с данными базы знаний Human Protein Atlas (версия 22.0) для экспрессии мРНК SOWAHD в линиях клеток NB4, HL-60, HepG2, HEK293, HeLa и K562.

Согласно транскриптомным данным, можно предположить, что мРНК SOWAHD экспрессируется преимущественно в клетках миелоидного происхождения. Экспрессия наблюдается в клетках острого миелоидного лейкоза, но отсутствует в клетках хронического миелогенного лейкоза. Однако с учётом ограничений для количественной оценки экспрессии транскриптов на основании величин TPM [25], а также известных на сегодняшний день фактов вариабельности омиксных данных даже в пределах одной клеточной линии [26],

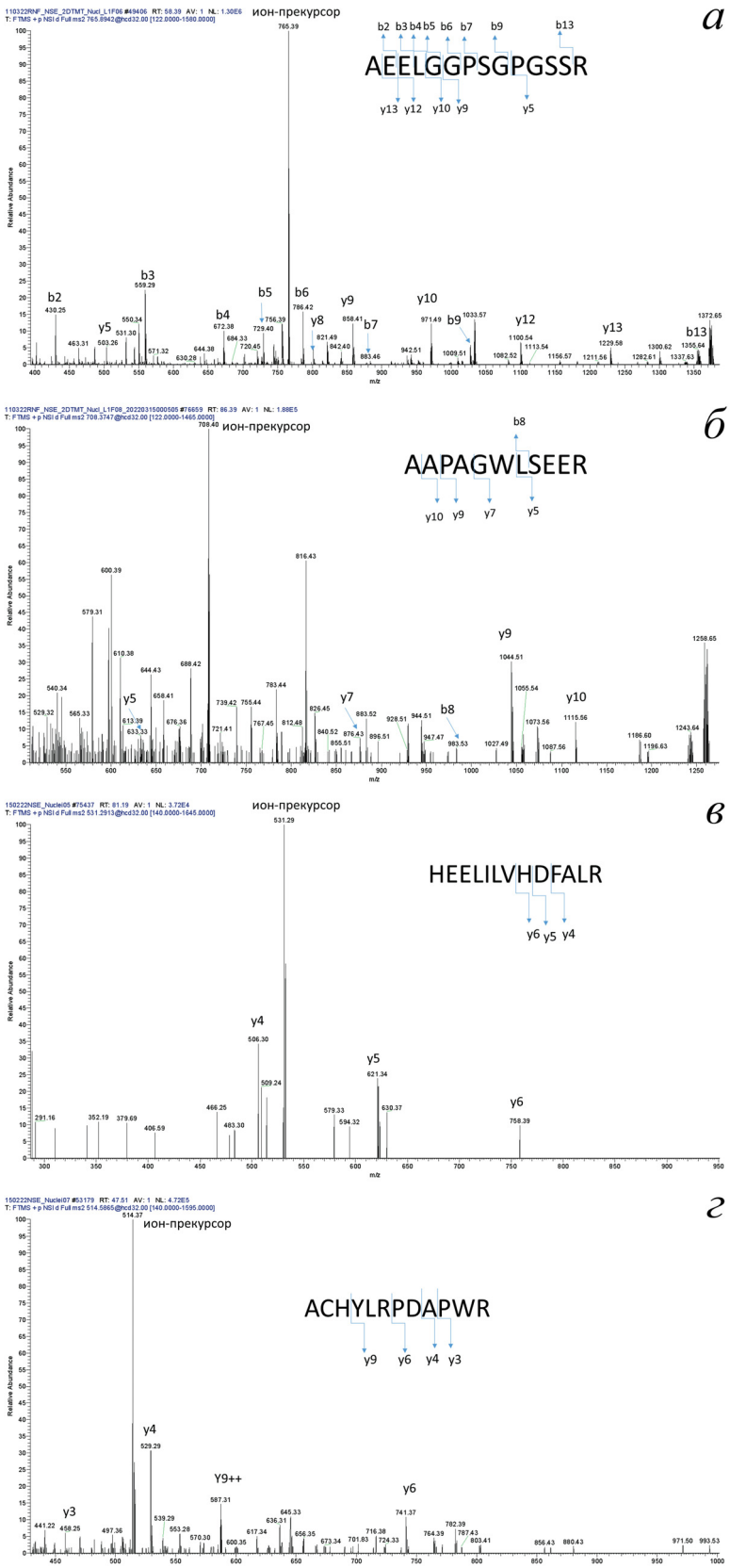

Рис. 4. MS2-спектры высокого разрешения, полученные для уникальных пептидов AEELGGPSGPGSSR (*а*), AAPAGWLSEER (*б*), HEELILVHDFALR (*в*) и ACHYLRPDAPWR (*г*), картируемых на белок SOWAHD. Спектры были визуализированы в программном обеспечении Thermo Xcalibur Qual Browser (Thermo Xcalibur 3.0.63), экспортированы в виде рисунка, после чего y- и b-фрагментарные ионы были аннотированы в ручном режиме. Пептиды были идентифицированы в автоматическом режиме в ПО MaxQuant. Номер MS2-спектра высокого разрешения с наилучшим качеством и соответствующий исходный файл с данными были указаны в файле «evidence.txt» выдачи MaxQuant

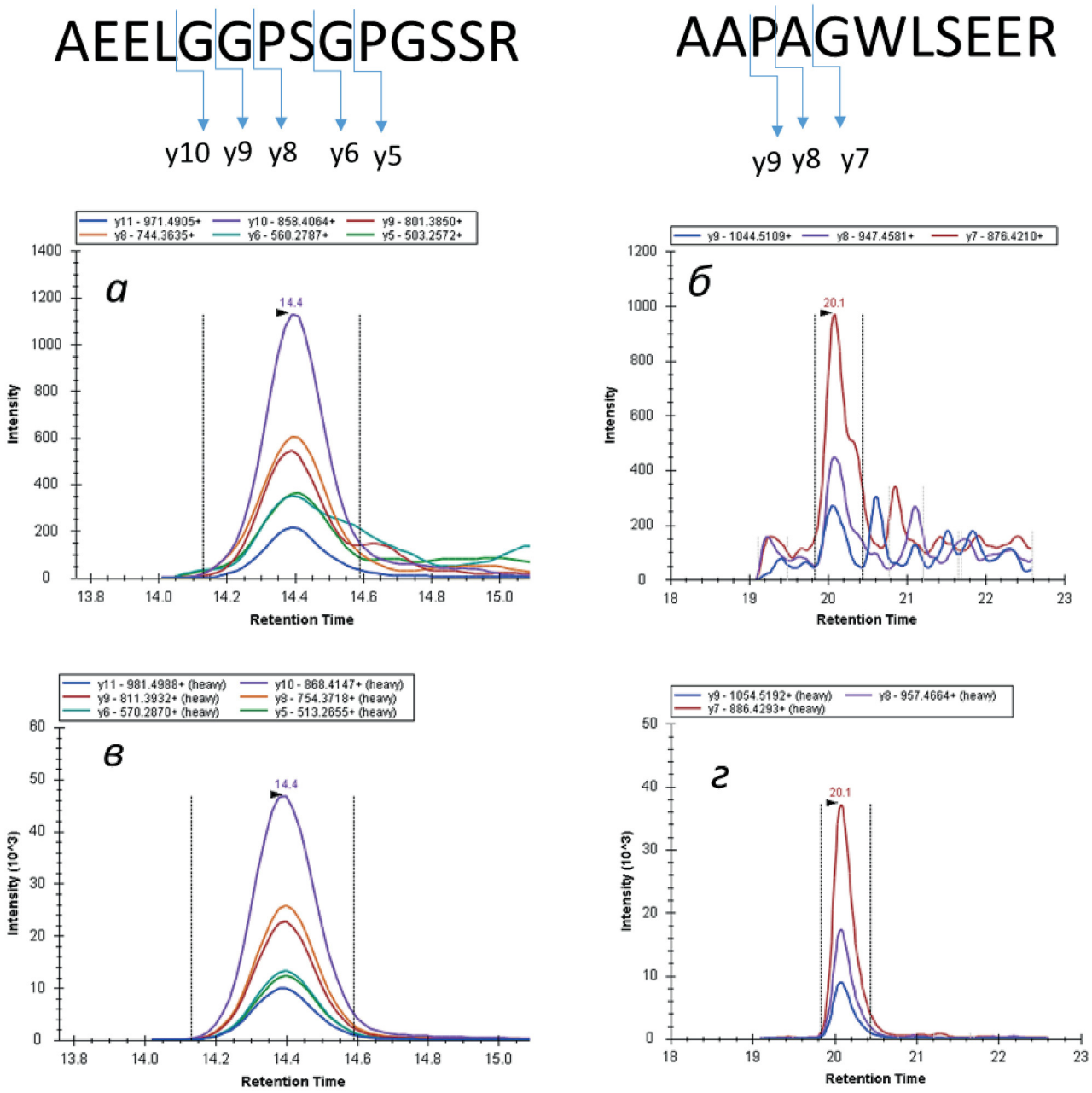

Рис. 5. Результат направленного масс-спектрометрического анализа в режиме SRM с использованием SIS AEELGGPSGPGSSR и AAPAGWLSEER. Приведены хроматограммы экстрагированного ионного тока для фрагментарных ионов нативных пептидов (*а* и *б*) и их синтетических аналогов (*в* и *г*)

тканеспецифичность экспрессии SOWAHD требует дальнейшей валидации с помощью ортогональных методов.

В данной работе фокус был направлен на исследование белкового продукта гена *SOWAHD*, а транскриптомные данные выступали как дополнительное доказательство экспрессии гена.

Динамическая экспрессия белка SOWAHD в ядрах клеток линии HL-60 в ходе ATRA-индуцированной гранулоцитарной дифференцировки. По результатам масс-спектрометрического анализа высокого разрешения ядер клеток линии HL-60 под действием ATRA белок SOWAHD был идентифицирован по 4-м уникальным пептидам (ACHYLRPDAPWR, HEELILVHDFALR, AAPAGWLSEER и AEELGGPSGPGSSR). Спектры высокого разрешения показаны на рис. 4.

Рис. 4 показывает, что масс-спектрометрический анализ высокого разрешения позволил идентифицировать уникальные пептиды (FDR < 0,01), соответствующие 15,6% аминокислотной последовательности белка SOWAHD.

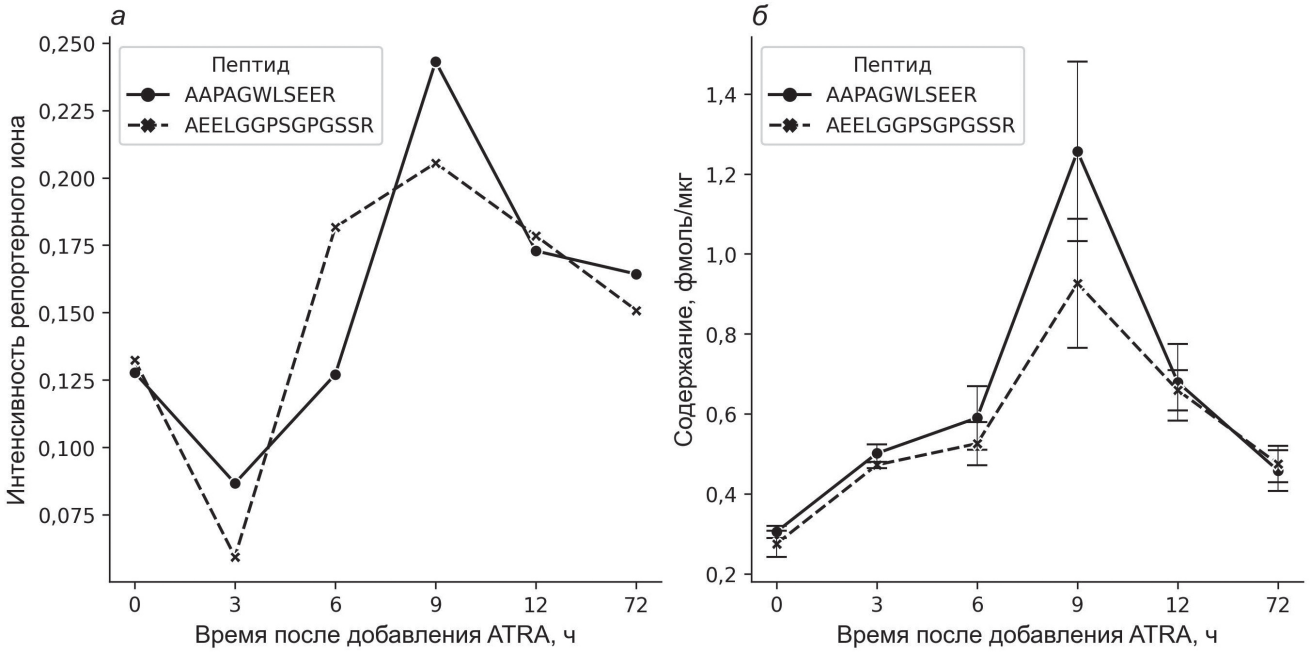

Рис. 6. Динамика экспрессии белка SOWAHD в ядерной фракции клеток линии HL-60 через 0, 3, 6, 9, 12 и 24 ч после добавления ATRA, индуцирующей гранулоцитарную дифференцировку. *а* – Результаты относительного количественного масс-спектрометрического анализа высокого разрешения с использованием TMT-меток, по оси *x* отложены временные точки после добавления ATRA, по оси *y* показана интенсивность репортерного TMT-иона. Интенсивность репортерного TMT-иона (MS2-спектр, спектр ионов-фрагментов) для пептидов AAPAGWLSEER и AEELGGPSGPGSSR была рассчитана в одном техническом повторе. *б* – Результаты количественного направленного масс-спектрометрического анализа с использованием изотопно-меченных пептидных стандартов (SRM/SIS), по оси *x* отложены временные точки после добавления ATRA, по оси y показано содержание протеотипических пептидов белка SOWAHD в фмоль/мкг общего белка. Данные получены для уникальных протеотипических пептидов AAPAGWLSEER и AEELGGPSGPGSSR. SRM/SIS-измерения были осуществлены в 3-х технических повторах в образцах ядер клеток HL60, индуцированных к дифференцировке ATRA в 3-х биологических повторах; всего на каждую временную точку приходится 9 LC/SRM-измерений

Для подтверждения идентификации белка SOWAHD были синтезированы SIS с последовательностями AAPAGWLSEER и AEELGGPSGPGSSR. Изотопно-меченные и соответствующие нативные пептиды были зарегистрированы методом направленного масс-спектрометрического анализа в режиме мониторинга выбранных реакций в клетках линии HL-60 (рис. 5).

Рис. 5, *б* показывает более высокий уровень шума для сигнала от пептида AAPAGWLSEER по сравнению с пептидом AEELGGPSGPGSSR. Низкая концентрация целевого белка, а также тот факт, что в этой области градиента может элюировать больше интерферирующих пептидов, может объяснять высокий уровень шума. С другой стороны, время элюирования и паттерн фрагментации полностью совпадают для нативного пептида и его синтетического аналога, за счёт чего мы считаем его измерение достоверным. Кроме того, профили экспрессии, а также абсолютное содержание в различные временные точки после добавления ATRA сходны для двух целевых пептидов, картируемых на белок SOWAHD.

Уникальные пептиды AAPAGWLSEER и AEELGGPSGPGSSR позволили осуществить количественный анализ белка SOWAHD с использованием TMT-меток (рис. 6, *а*), а также с использованием направленного масс-спектрометрического метода SRM/SIS (рис. 6, *б*) в ядерной фракции клеток линии HL-60 через 3, 6, 9, 12 и 72 ч после добавления ATRA.

Масс-спектрометрический анализ высокого разрешения с TMT-метками показал увеличение содержания белка SOWAHD в ядерной фракции в 1,9 раз через 9 ч после добавления ATRA в среду культивирования клеток. По результатам SRM/SIS-анализа содержание белка SOWAHD было определено на уровне 0,3 ± ± 0,06, 0,5 ± 0,06, 0,59 ± 0,2, 1,25 ± 0,63, 0,67 ± ± 0,28 и 0,45 ± 0,15 фмоль/мкг по пептиду AAPAGWLSEER, а также на уровне 0,27 ± ± 0,07, 0,47 ± 0,02, 0,52 ± 0,16, 0,92 ± 0,45, 0,63 ± ± 0,17 и 0,47 ± 0,13 фмоль/мкг по пептиду AEELGGPSGPGSSR через 0, 3, 6, 9, 12 и 72 ч после добавления ATRA соответственно. Содержание белка SOWAHD в ядерной фракции значимо (p-value < 0,01) увеличивалось в 1,6, 1,9, 3,7, 2,3 и 1,5 раза через 3, 6, 9, 12 и 72 ч

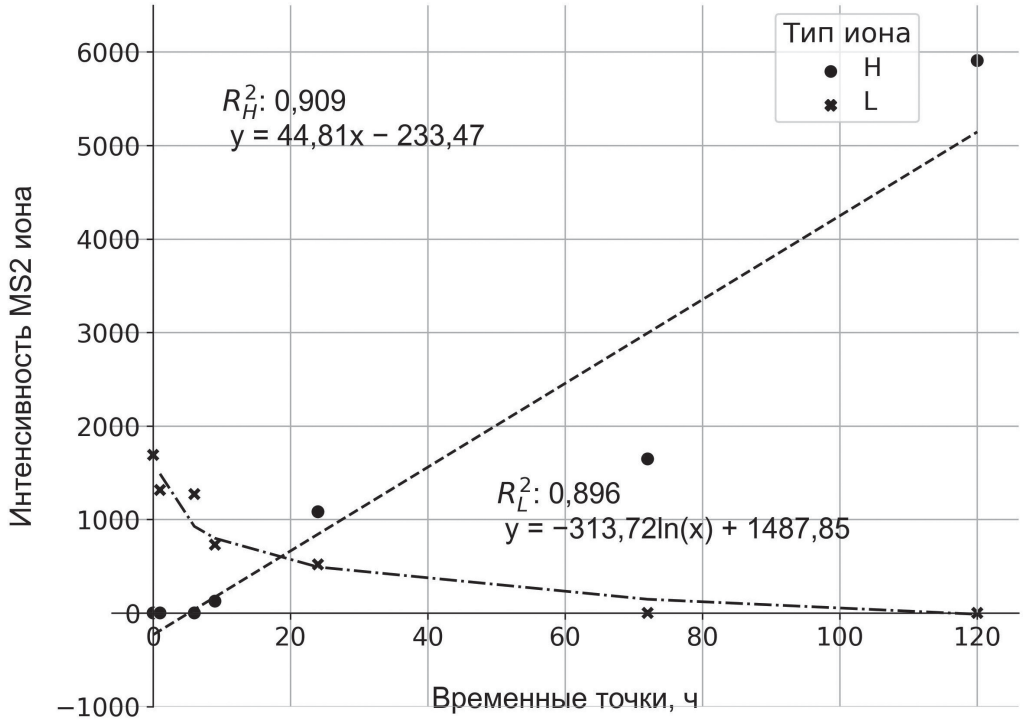

Рис. 7. Динамика синтеза и деградации белкового продукта гена *SOWAHD*, определённая с использованием пульсового SILAC-мечения в комбинации с направленным масс-спектрометрическим анализом SRM. Графики зависимости площади под пиком от времени культивирования в среде SILAC для «лёгкой» (L) и «тяжёлой» (H) формы пептида SOWAHD. По оси *x* отложены временные точки (0, 1, 6, 9, 24, 72 и 120 ч), по оси *y* – площадь под пиком для наиболее интенсивного дочернего иона (MS2, фрагмент y10+, −GGPSGPGSSR) «лёгкой» (L) и «тяжёлой» (H) формы пептида SOWAHD (AEELGGPSGPGSSR). Графики аппроксимирующих функций показаны штриховой (- -) линией (линейная зависимость) и штрихпунктирной (-.) линией (логарифмическая зависимость) для «тяжёлой» (H) и «легкой» (L) формы пептида SOWAHD соответственно. Представлены результаты для первого технического повтора, для второго и третьего технического повтора графики приведены на рис. 2–3 П Приложения 2

после добавления индуктора гранулоцитарной дифференцировки соответственно. Рис. 6 показывает, что профили экспрессии, полученные для пептидов AAPAGWLSEER и AEELGGPSGPGSSR, оказались похожими между собой.

В силу низкого содержания белка SOWAHD, а также вероятностного характера фрагментации в ходе панорамного масс-спектрометрического анализа спектры фрагментации (MS2) в качестве, достаточном для вычисления интенсивности репортерных TMT-ионов, были получены только в одном техническом повторе для пептидов AAPAGWLSEER и AEELGGPSGPGSSR. Тем не менее наблюдалась сходимость результатов, полученных с помощью различных методов количественной оценки – масс-спектрометрии высокого разрешения с TMT-метками и направленной масс-спектрометрии в режиме SRM.

Динамика синтеза и деградации белкового продукта SOWAHD. Время полураспада белкового продукта SOWAHD. Для определения динамики синтеза и деградации белкового продукта SOWAHD применили комбинацию пульсового SILAC-мечения и направленного масс-спектрометрического анализа SRM. Принцип пульсового мечения SILAC описан в разделе «Материалы и методы». Динамику встраивания ^{13}C-меченного лизина или аргинина из среды культивирования в белок SOWAHD определяли по результатам SRM-измерений «лёгкой» (L) и «тяжёлой» (H) формы пептида SOWAHD (AEELGGPSGPGSSR), как показано на рис. 7.

Как показывает рис. 7, направленный масс-спектрометрический анализ позволил определить аппроксимирующие функции, описывающие динамику синтеза и деградации SOWAHD, на основании площади под пиком для наиболее интенсивного дочернего иона пептида AEELGGPSGPGSSR (MS2, фрагмент y10+, −GGPSGPGSSR). Измерения осуществлялись для пептида AEELGGPSGPGSSR, поскольку для него было зарегистрировано больше фрагментов, как по данным масс-спектрометрии высокого разрешения (рис. 4, *а*), так и по результатам направленного анализа (рис. 5). Время полураспада белкового продукта SOWAHD ($t_{1/2}$) было рассчитано в трёх технических повторах и составило 19 ± 7 ч.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Ген *SOWAHD* относится к семейству *SOWAH*, в которое входят ещё 3 гена: *SOWAHA*, *SOWAHB* и *SOWAHC*. Согласно нашим панорамным протеомным данным, только продукт гена *SOWAHD* был зарегистрирован на уровне белка в клетках линии HL-60. В то же время наши экспериментальные транскриптомные данные показали, что в линиях эпителиального происхождения (HeLa, HepG2 и HEK293) наибольшая экспрессия мРНК, на порядок превышающая таковую для остальных генов семейства, наблюдалась для гена *SOWAHC*, в то время как в миелоидных линиях (HL-60 и NB4) превалировала экспрессия *SOWAHD* на фоне пониженной экспрессии других генов семейства.

Данные об экспрессии белкового продукта SOWAHD отличаются в основных протеомных базах знаний Uniprot, NextProt и Human Protein Atlas. Так, в базе знаний Uniprot указана информация об экспрессии продукта гена *SOWAHD* на уровне транскрипта (https://www.uniprot.org/uniprotkb/A6NJG2/entry) (PE2). В базе знаний Human Protein Atlas, опирающейся на анализ белковых продуктов с помощью моноклональных антител, указано, что экспрессия гена *SOWAHD* на уровне белка подтверждена (PE1), но внутриклеточная локализация неизвестна, а информация об уровне экспрессии продукта этого гена приведена только на уровне мРНК (https://www.proteinatlas.org/ENSG00000187808-SOWAHD). Наконец, база знаний NextProt, компилирующая масс-спектрометрические данные, категоризирует продукт гена *SOWAHD* как экспрессирующийся на уровне белка (PE1), однако информация о его внутриклеточной локализации, биологической функции и медицинской значимости отсутствует.

По материалам базы данных NextProt, в семействе SOWAH наиболее исследованным является белок SOWAHC, для которого была установлена цитозольная локализация, а также функциональная вовлечённость в сигнальный путь цикла RHOB GTPазы. Для всех генов семейства, кроме *SOWAHD*, также приведены значимые мутации, которые могут определять медицинскую значимость. В контексте онкологических заболеваний было показано, что гипометилирование гена *SOWAHC* ассоциировано с неблагоприятным прогнозом для пациентов с раком лёгких [27]. Высокая экспрессия гена *SOWAHB* на уровне мРНК является благоприятным прогностическим фактором для пациентов с почечной светлоклеточной карциномой [28]. Экспрессия гена *SOWAHA* представляет собой благоприятный прогностический фактор в случае рака поджелудочной железы [29]. Исследования были выполнены на основе биоинформатического метаанализа данных, полученных из ресурса The Cancer Genome Atlas, и только в случае с геном *SOWAHA* была проведена валидация результатов с помощью транскриптомного анализа.

Как показали наши протеомные данные, экспрессия белка SOWAHD, измеренная в ядре клеток HL-60, увеличивалась в процессе ATRA-индуцированной гранулоцитарной дифференцировки, что позволяет предположить ядерную локализацию белка и его значимость для созревания нейтрофилов. Пик увеличения экспрессии белка приходился на 9 ч после добавления индуктора с последующим снижением в более поздние временные точки, что может указывать на функциональную значимость белка SOWAHD для ранних стадий индуцированной гранулоцитарной дифференцировки.

Ранее с помощью направленной масс-спектрометрии в ядрах клеток линии HL-60 были зарегистрированы транскрипционные факторы (ТФ) HIC1, STAT1 и RBPJ на уровне ~0,5–1 фмоль/мкг общего белка [13]. Количество белка SOWAHD, измеренное в ядерной фракции в данной работе, схоже с содержанием вышеупомянутых ТФ.

В данной работе, исследуя динамику белка SOWAHD в процессе ATRA-индуцированной гранулоцитарной дифференцировки, выявили дугообразный профиль экспрессии с увеличением содержания в ранние временные точки и возвращением уровня экспрессии к базовой линии через 72 ч после обработки ATRA. На основании данных направленной масс-спектрометрии похожий профиль был получен для белков, вовлечённых в регуляцию апоптоза и репарации ДНК (STAT1, CASP3, PARP1, PRKDC и UBE2I) [13]. Анализ данных панорамной масс-спектрометрии, полученных нами ранее [13], также выявил «дугообразный» профиль экспрессии для белков, вовлечённых в альтернативный сплайсинг и репликацию ДНК (рис. 8). Кроме того, данные белки, согласно анализу относительно базы данных TRRUST, могут регулироваться транскрипционными факторами MYCN и SF1 (рис. 8).

Скорость полураспада белка SOWAHD (~19 ч) сходна с таковой для сигнальных и регуляторных белков [34]. Нужно отметить, что в данном исследовании пульсовое мечение SILAC применяли совместно с направленным масс-спектрометрическим анализом в режиме

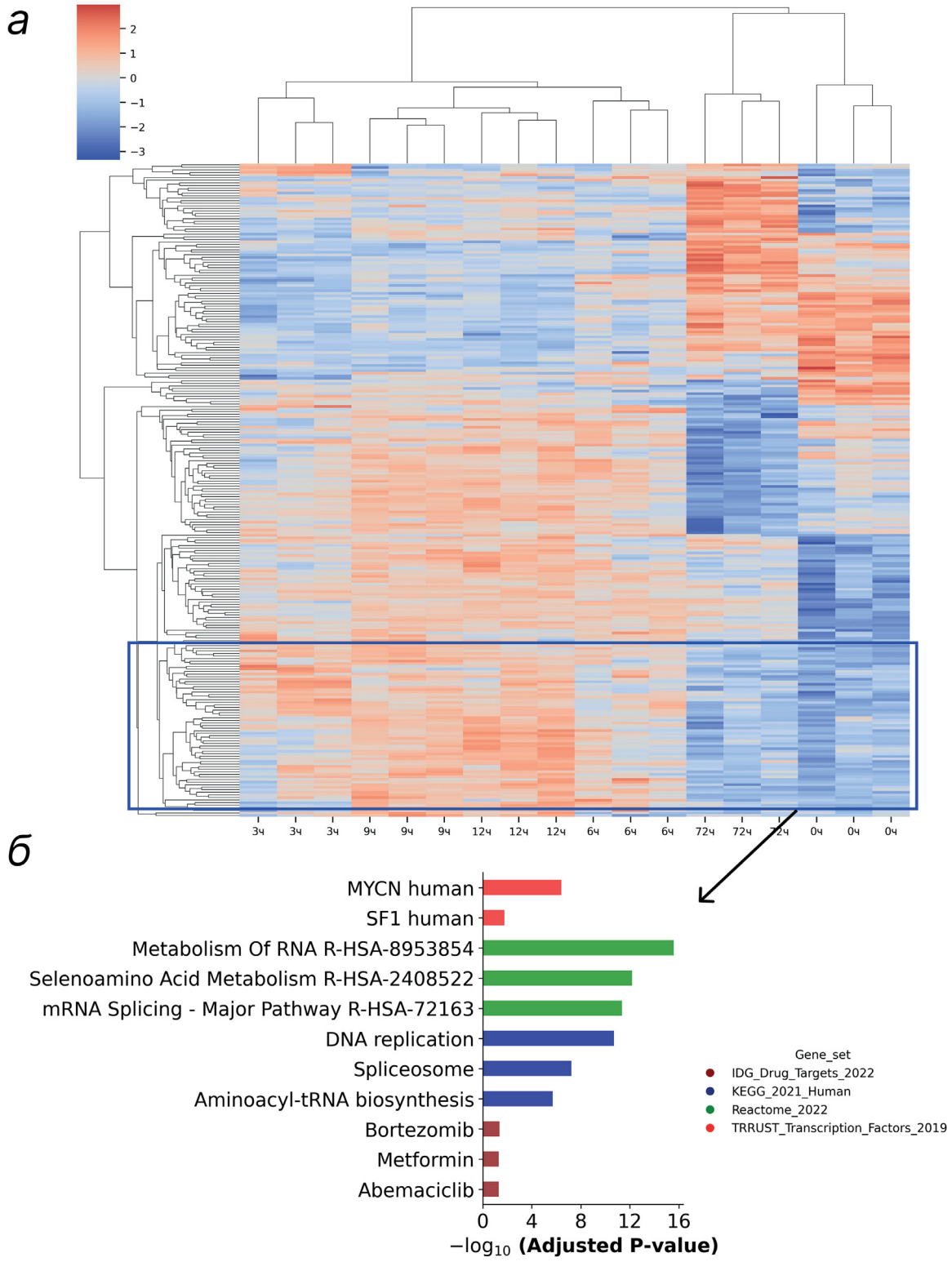

Рис. 8. Результаты анализа данных панорамной масс-спектрометрии. *а* – «Тепловая» карта, демонстрирующая профиль экспрессии 254 белков, содержание которых в ядерной фракции клеток линии HL-60 значимо (S0 = 0,1, FDR = 0,01) изменялось в процессе ATRA-индуцированной гранулоцитарной дифференцировки. Для анализа были использованы масс-спектрометрические данные, полученные ранее [13]. Для оценки дифференциальной экспрессии применили дисперсионный анализ ANOVA с поправкой на множественность сравнений с помощью пермутационного теста. *б* – Для кластера, компоненты которого демонстрировали сходную с SOWAHD экспрессию (выделен синим), выполнена аннотация относительно сигнальных путей (Reactome [30] и KEGG [31]), наличия потенциальной транскрипционной регуляции (TRRUST [32]), возможной чувствительности к фармакологическим препаратам (IDG [33]). Можно предположить, что белок SOWAHD может быть вовлечён в один или несколько из вышеупомянутых биологических процессов, однако эта гипотеза нуждается в дальнейшей экспериментальной проверке

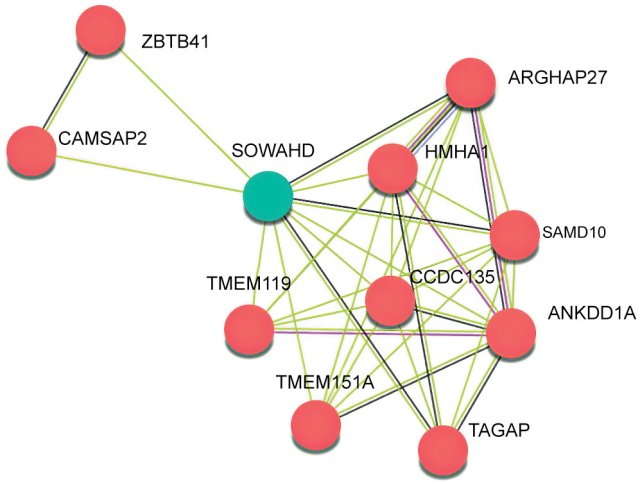

Рис. 9. Схема белок-белковых взаимодействий, вовлечённых в васкуляризацию и гемопоэз у рыб вида *Danio rerio* (по материалам базы данных STRING [36]), включающая белок-ортолог SOWAHD

SRM. Панорамный анализ, обычно применяемый в пульсовом мечении SILAC, не всегда позволяет вычислить величину площади под пиком для прекурсорных ионов, что приводит к наличию пропущенных значений (missing value) в количественных SILAC-данных. Особенно остро эта проблема стоит для низкокопийных белков. Использование направленной масс-спектрометрии позволяет улучшить точность и чувствительность определения скорости полураспада [21].

Существуют гипотезы о выполняемой функции для ортологов гена *SOWAHD* у других видов позвоночных. В экспериментах на рыбах вида *Danio rerio* было показано, что ортолог гена *SOWAHD* (*ankrd58*) вовлечён в процесс васкуляризации и гемопоэз [35] (рис. 9).

С учётом экспрессии мРНК и белка SOWAHD в миелоидных клетках он может также быть задействован в гемопоэзе у человека.

На основании сходства белковой последовательности можно заключить, что в структуре белка SOWAHD должны присутствовать анкириновые повторы (https://www.uniprot.org/uniprotkb/A6NJG2/entry). Анкириновые повторы являются наиболее распространёнными мотивами, вовлечёнными в белок-белковые взаимодействия. Белки, содержащие анкириновые повторы, такие как опухолевый супрессор p16, рецептор Nank сигнального пути Notch, ингибитор IκBα ядерного фактора каппа B и т. д., вовлечены в регуляцию клеточного цикла, активацию транскрипции, апоптоза и пролиферации у различных биологических видов, в том числе и у человека [37, 38].

Настоящее исследование было посвящено всестороннему протеомному анализу белка SOWAHD, при этом белковый продукт гена *KIAA1143* также может быть потенциальным регулятором гранулоцитарной дифференцировки. В будущем планируется провести протеомный анализ белка KIAA1143 для углублённого представления о процессах индуцированного созревания лейкозных клеток.

ЗАКЛЮЧЕНИЕ

Данные протеомного анализа указывают, что SOWAHD вовлечён в процесс ATRA-индуцированной гранулоцитарной дифференцировки. Планируемые эксперименты по нокдауну SOWAHD будут призваны проверить эту теорию. В протеомной базе данных NextProt хранится информация об экспрессии белка SOWAHD, собранная на основании множества экспериментов, однако систематическое исследование экспрессии продукта белок-кодирующего гена SOWAHD на уровне мРНК и белка на одном и том же материале проводилось впервые. Также уникальными являются данные по времени полураспада ($t_{1/2}$) белкового продукта SOWAHD. Для определения величины $t_{1/2}$ был применён направленный масс-спектрометрический метод, который позволяет анализировать белки с низким содержанием.

Вклад авторов. С.Е. Новикова, А.Л. Русанов, В.Г. Згода – концепция и руководство работой; С.Е. Новикова, Т.В. Толстова, Н.А. Соловьева, Т.Е. Фарафонова, О.В. Тихонова, Л.К. Курбатов – проведение экспериментов; С.Е. Новикова, О.В. Тихонова, Л.К. Курбатов, А.Л. Русанов, В.Г. Згода – обсуждение результатов исследования; С.Е. Новикова – написание текста; В.Г. Згода – редактирование текста статьи.

Финансирование. Работа выполнена при финансовой поддержке Российского научного фонда (грант № 21-74-20122).

Благодарности. Исследование выполнено с использованием оборудования Центра коллективного пользования «Протеом человека» в ИБМХ.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Соблюдение этических норм. Настоящая статья не содержит описания каких-либо исследований с участием людей или животных в качестве объектов.

Дополнительные материалы. Приложение к статье опубликовано на сайте журнала «Биохимия» (https://biochemistrymoscow.com).

СПИСОК ЛИТЕРАТУРЫ

1. Consortium UniProt (2023) UniProt: The Universal Protein Knowledgebase in 2023, *Nucleic Acids Res.*, **51**, D523-D531, doi: 10.1093/nar/gkac1052.
2. Lane, L., Argoud-Puy, G., Britan, A., Cusin, I., Duek, P. D., Evalet, O., Gateau, A., Gaudet, P., Gleizes, A., Masselot, A., et al. (2012) NeXtProt: a knowledge platform for human proteins, *Nucleic Acids Res.*, **40**, D76-83, doi: 10.1093/nar/gkr1179.
3. Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015) Proteomics. Tissue-based map of the human proteome, *Science*, **347**, 1260419, doi: 10.1126/science.1260419.
4. Kim, M. S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., Madugundu, A. K., Kelkar, D. S., Isserlin, R., Jain, S., et al. (2014) A draft map of the human proteome, *Nature*, **509**, 575-581, doi: 10.1038/nature13302.
5. Adhikari, S., Nice, E. C., Deutsch, E. W., Lane, L., Omenn, G. S., Pennington, S. R., Paik, Y. K., Overall, C. M., Corrales, F. J., Cristea, I. M., et al. (2020) A high-stringency blueprint of the human proteome, *Nat. Commun.*, **11**, 5301, doi: 10.1038/s41467-020-19045-9.
6. Salzberg, S. L. (2018) Open questions: how many genes do we have? *BMC Biol.*, **16**, 94, doi: 10.1186/s12915-018-0564-x.
7. Kopylov, A. T., Ponomarenko, E. A., Ilgisonis, E. V., Pyatnitskiy, M. A., Lisitsa, A. V., Poverennaya, E. V., Kiseleva, O. I., Farafonova, T. E., Tikhonova, O. V., Zavialova, M. G., et al. (2019) 200+ protein concentrations in healthy human blood plasma: targeted quantitative SRM SIS screening of chromosomes 18, 13, Y, and the mitochondrial chromosome encoded proteome, *J. Proteome Res.*, **18**, 120-129, doi: 10.1021/acs.jproteome.8b00391.
8. Poverennaya, E., Kiseleva, O., Ilgisonis, E., Novikova, S., Kopylov, A., Ivanov, Y., Kononikhin, A., Gorshkov, M., Kushlinskii, N., Archakov, A., et al. (2020) Is it possible to find needles in a haystack? Meta-analysis of 1000+ MS/MS files provided by the Russian proteomic consortium for mining missing proteins, *Proteomes*, **8**, 12, doi: 10.3390/PROTEOMES8020012.
9. Paik, Y. K., Overall, C. M., Corrales, F., Deutsch, E. W., Lane, L., and Omenn, G. S. (2018) toward completion of the human proteome parts list: progress uncovering proteins that are missing or have unknown function and developing analytical methods, *J. Proteome Res.*, **17**, 4023-4030, doi: 10.1021/acs.jproteome.8b00885.
10. Thul, P. J., and Lindskog, C. (2018) The human protein atlas: a spatial map of the human proteome, *Protein Sci.*, **27**, 233-244, doi: 10.1002/pro.3307.
11. Van Bortle, K., and Corces, V. G. (2012) Nuclear organization and genome function, *Annu. Rev. Cell Dev. Biol.*, **28**, 163-187, doi: 10.1146/annurev-cellbio-101011-155824.
12. Vakhrushev, I. V., Novikova, S. E., Tsvetkova, A. V., Karalkin, P. A., Pyatnitskii, M. A., Zgoda, V. G., and Yarygin, K. N. (2018) Proteomic profiling of HL-60 cells during ATRA-induced differentiation, *Bull. Exp. Biol. Med.*, **165**, 530-543, doi: 10.1007/s10517-018-4210-y.
13. Novikova, S., Tolstova, T., Kurbatov, L., Farafonova, T., Tikhonova, O., Soloveva, N., Rusanov, A., Archakov, A., and Zgoda, V. (2022) Nuclear proteomics of induced leukemia cell differentiation, *Cells*, **11**, 3221, doi: 10.3390/cells11203221.
14. Zheng, P. Z., Wang, K. K., Zhang, Q. Y., Huang, Q. H., Du, Y. Z., Zhang, Q. H., Xiao, D. K., Shen, S. H., Imbeaud, S., Eveno, E., et al. (2005) Systems analysis of transcriptome and proteome in retinoic acid/arsenic trioxide-induced cell differentiation apoptosis of promyelocytic leukemia, *Proc. Natl. Acad. Sci. USA*, **102**, 7653-7658, doi: 10.1073/pnas.0502825102.
15. Wang, W.-J., Tang, W., and Qiu, Z.-Y. (2009) Comparative proteomics analysis on differentiation of human promyelocytic leukemia HL-60 cells into granulocyte and monocyte lineages, *Chinese J. Cancer*, **28**, 117-121.
16. Novikova, S., Tikhonova, O., Kurbatov, L., Farafonova, T., Vakhrushev, I., Lupatov, A., Yarygin, K., and Zgoda, V. (2021) Omics technologies to decipher regulatory networks in granulocytic cell differentiation, *Biomolecules*, **11**, 907, doi: 10.3390/biom11060907.
17. Jian, P., Li, Z. W., Fang, T. Y., Jian, W., Zhuan, Z., Mei, L. X., Yan, W. S., and Jian, N. (2011) Retinoic acid induces HL-60 cell differentiation via the upregulation of MiR-663, *J. Hematol. Oncol.*, **4**, 20, doi: 10.1186/1756-8722-4-20.
18. Schwanhäusser, B., Gossen, M., Dittmar, G., and Selbach, M. (2009) Global analysis of cellular protein translation by pulsed SILAC, *Proteomics*, **9**, 205-209, doi: 10.1002/pmic.200800275.
19. Claydon, A. J., and Beynon, R. (2012) Proteome dynamics: revisiting turnover with a global perspective, *Mol. Cell. Proteomics*, **11**, 1551-1565, doi: 10.1074/mcp.O112.022186.
20. Ross, A. B., Langer, J. D., and Jovanovic, M. (2021) Proteome turnover in the spotlight: approaches, applications, and perspectives, *Mol. Cell. Proteomics*, **20**, doi: 10.1074/mcp.R120.002190.
21. Holman, S. W., Hammond, D. E., Simpson, D. M., Waters, J., Hurst, J. L., and Beynon, R. J. (2016) Protein turnover measurement using selected reaction monitoring-mass spectrometry (SRM-MS), *Philos.*

Trans. R. Soc. A Math. Phys. Eng. Sci., **374**, 20150362, doi: 10.1098/rsta.2015.0362.
22. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression, *Nat. Methods*, **14**, 417-419, doi: 10.1038/nmeth.4197.
23. Novikova, S. E., Vakhrushev, I. V., Tsvetkova, A. V., Shushkova, N. A., Farafonova, T. E., Yarygin, K. N., and Zgoda, V. G. (2019) Proteomics of transcription factors: identification of pool of HL-60 cell line-specific regulatory proteins [in Russian], *Biomed. Khim.*, **65**, doi: 10.18097/PBMC20196504294.
24. Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis, *Nat. Methods*, **6**, 359-362, doi: 10.1038/nmeth.1322.
25. Mohammad, N. S., Nazli, R., Zafar, H., and Fatima, S. (2022) Effects of lipid based multiple micronutrients supplement on the birth outcome of underweight pre-eclamptic women: a randomized clinical trial, *Pak. J. Med. Sci.*, **38**, 219-226, doi: 10.12669/pjms.38.1.4396.
26. Liu, Y., Mi, Y., Mueller, T., Kreibich, S., Williams, E. G., Van Drogen, A., Borel, C., Frank, M., Germain, P.-L., Bludau, I., et al. (2019) Multi-omic measurements of heterogeneity in HeLa cells across laboratories, *Nat. Biotechnol.*, **37**, 314-322, doi: 10.1038/s41587-019-0037-y.
27. Zhu, Q., Wang, J., Zhang, Q., Wang, F., Fang, L., Song, B., Xie, C., and Liu, J. (2020) Methylation-driven Genes PMPCAP1, SOWAHC and ZNF454 as potential prognostic biomarkers in lung squamous cell carcinoma, *Mol. Med. Rep.*, **21**, 1285-1295, doi: 10.3892/mmr.2020.10933.
28. Ruan, B., Feng, X., Chen, X., Dong, Z., Wang, Q., Xu, K., Tian, J., Liu, J., Chen, Z., Shi, W., et al. (2020) Identification of a set of genes improving survival prediction in kidney renal clear cell carcinoma through integrative reanalysis of transcriptomic data, *Dis. Markers*, **2020**, 8824717, doi: 10.1155/2020/8824717.
29. Mori, Y., Yokota, H., Hoshino, I., Iwatate, Y., Wakamatsu, K., Uno, T., and Suyari, H. (2021) Deep learning-based gene selection in comprehensive gene analysis in pancreatic cancer, *Sci. Rep.*, **11**, 16521, doi: 10.1038/s41598-021-95969-6.
30. Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., Griss, J., Sevilla, C., Matthews, L., Gong, C., et al. (2022) The reactome pathway knowledgebase, *Nucleic Acids Res.*, **50**, D687-D692, doi: 10.1093/nar/gkab1028.
31. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation, *Nucleic Acids Res.*, **44**, D457-62, doi: 10.1093/nar/gkv1070.
32. Han, H., Shim, H., Shin, D., Shim, J. E., Ko, Y., Shin, J., Kim, H., Cho, A., Kim, E., Lee, T., et al. (2015) TRRUST: a reference database of human transcriptional regulatory interactions, *Sci. Rep.*, **5**, 11432, doi: 10.1038/srep11432.
33. Lin, Y., Mehta, S., Küçük-McGinty, H., Turner, J. P., Vidovic, D., Forlin, M., Koleti, A., Nguyen, D.-T., Jensen, L. J., Guha, R., et al. (2017) Drug target ontology to classify and integrate drug discovery data, *J. Biomed. Semantics*, **8**, 50, doi: 10.1186/s13326-017-0161-x.
34. Mathieson, T., Franken, H., Kosinski, J., Kurzawa, N., Zinn, N., Sweetman, G., Poeckel, D., Ratnu, V. S., Schramm, M., Becher, I., et al. (2018) Systematic analysis of protein turnover in primary cells, *Nat. Commun.*, **9**, 689, doi: 10.1038/s41467-018-03106-1.
35. Gomez, G., Lee, J. H., Veldman, M. B., Lu, J., Xiao, X., and Lin, S. (2012) Identification of vascular and hematopoietic genes downstream of *etsrp* by deep sequencing in zebrafish, *PLoS One*, **7**, e31658, doi: 10.1371/journal.pone.0031658.
36. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., et al. (2023) The STRING database in 2023: protein-protein association networks and functional Enrichment analyses for any sequenced genome of interest, *Nucleic Acids Res.*, **51**, D638-D646, doi: 10.1093/nar/gkac1000.
37. Kang, Y., Xie, H., and Zhao, C. (2019) Ankrd45 is a novel ankyrin repeat protein required for cell proliferation, *Genes (Basel)*, **10**, 462, doi: 10.3390/genes10060462.
38. Kumar, A., and Balbach, J. (2021) Folding and stability of ankyrin repeats control biological protein function, *Biomolecules*, **11**, 840, doi: 10.3390/biom11060840.

PROTEOMIC APPROACH TO INVESTIGATE THE EXPRESSION, LOCALIZATION, AND FUNCTIONS OF THE PROTEIN PRODUCT OF THE *SOWAHD* GENE DURING GRANULOCYTIC DIFFERENTIATION

S. E. Novikova*, T. V. Tolstova, N. A. Soloveva*, T. E. Farafonova, O. V. Tikhonova, L. K. Kurbatov, A. L. Rusanov, and V. G. Zgoda*

Institute of Biomedical Chemistry,
119121 Moscow, Russia; e-mail: novikova.s.e3101@gmail.com, victor.zgoda@gmail.com

Cataloging human proteins, determining their level of content, cellular localization, function performed, and potential medical significance are important tasks facing the global proteomic community. At present, the localization and functions of protein products for almost half of the protein-coding genes are unknown or poorly understood. The study of the organelle proteome is a promising approach to reveal the localization and functions of human proteins. The nuclear proteome is of particular interest because many nuclear-localized proteins, such as transcription factors, perform regulatory functions that determine cell fate. According to the results of a meta-analysis of the nuclear proteome, or nucleome, of HL-60 cells treated by all-*trans*-retinoic acid (ATRA), it was revealed that the function and localization of the protein product of the *SOWAHD* gene is poorly understood, in addition, there is no comprehensive information on the expression of *SOWAHD* at the protein level. In HL-60 cells for the protein-coding gene *SOWAHD*, mRNA expression was determined at the level of 6.4 ± 0.7 transcripts per million molecules. Using targeted mass spectrometry, the content of SOWAHD protein was measured in the range of 0.27-1.25 fmol/μg of total protein. Using stable isotope pulse-chase labeling, the half-life for the protein product of the *SOWAHD* gene was determined to be approximately 19 h. Proteomic profiling of the nuclear fraction of HL-60 cells showed that the SOWAHD protein content increased during ATRA-induced granulocytic differentiation, peaking at 9 h after the addition of the inductor and followed by a decrease at later time points. The results of the study indicate for the first time the nuclear localization and involvement of the protein product of the *SOWAHD* gene in induced granulocytic differentiation.

Keywords: protein product of the *SOWAHD* gene, transcriptome, proteome, ATRA, cell nucleus, mass spectrometry, pulse labeling with stable isotopes, dynamics of protein synthesis and degradation