

AgeMeta: КОЛИЧЕСТВЕННАЯ БАЗА ДАННЫХ ИЗМЕНЕНИЙ ЭКСПРЕССИИ ГЕНОВ В ПРОЦЕССЕ СТАРЕНИЯ МЛЕКОПИТАЮЩИХ

© 2024 С.А. Тихонов^{1,2}, М.А. Батин³, В.Н. Гладышев⁴,
С.Е. Дмитриев^{1,2}, А.Э. Тышковский^{1,4*}

¹ НИИ физико-химической биологии имени А.Н. Белозерского,
Московский государственный университет имени М.В. Ломоносова, 119992 Москва, Россия

² Московский государственный университет имени М.В. Ломоносова,
факультет биоинженерии и биоинформатики, 119234 Москва, Россия

³ Open Longevity, бульвар Вентура, Шерман-Окс, Калифорния, 91403 США

⁴ Гарвардская медицинская школа,
Женская больница Бригема, медицинский факультет, кафедра генетики,
Бостон, Массачусетс, 02115 США; электронная почта: atyshkovskii@bwh.harvard.edu

Поступила в редакцию 12.10.2023

После доработки 12.12.2023

Принята к публикации 13.02.2024

База данных AgeMeta содержит системное и количественное описание старения млекопитающих на уровне экспрессии генов. В ней представлены возрастные транскриптомные изменения в различных тканях человека, мыши и крысы на основе комплексного мета-анализа 122 общедоступных наборов данных по экспрессии генов из 26 исследований. AgeMeta содержит интуитивно понятный визуальный интерфейс для количественной оценки транскриптомики старения как на уровне отдельных генов, так и на уровне их функциональных групп, и позволяет легко сравнивать различные ткани и виды животных. Кроме того, все данные из AgeMeta могут быть скачаны пользователем и проанализированы независимо. Данная работа способствует лучшему пониманию сложной сети биологических процессов, лежащих в основе старения млекопитающих. AgeMeta находится в свободном доступе по ссылке: <https://age-meta.com/>.

КЛЮЧЕВЫЕ СЛОВА: старение, экспрессия генов, дифференциальная экспрессия, транскриптомика, млекопитающие, мета-анализ, AgeMeta, age-meta, база данных, GSEA.

DOI: 10.31857/S0320972524020099 EDN: XMHQEG

ПОСВЯЩЕНИЕ

Мы посвящаем эту работу светлой памяти Владимира Петровича Скулачёва – нашего уважаемого учителя, наставника, коллеги и выдающегося учёного. Академик Скулачёв известен в первую очередь своими прорывными работами в области клеточной биоэнергетики и функции митохондрий, однако в последние годы его неугасимый интерес к науке также распространялся на биологию старения, что привело к множеству выдающихся открытий. Наследие Владимира Петровича определяется не только его научными достижениями, но и тем неизгладимым впечатлением, которое он производил на своих коллег благодаря своей мудрости, доброте и замечательному характеру.

Его глубокое влияние простиралось далеко за пределы академических кругов, поскольку он активно занимался также организаторской и педагогической деятельностью. В наших сердцах он навсегда займёт особое место, память о нём будет служить источником вдохновения для нынешнего и будущих поколений.

ВВЕДЕНИЕ

Старение является основным фактором риска развития множества хронических заболеваний, что делает его значимой и глобальной медицинской проблемой [1]. Понимание молекулярных основ старения и межвидовых различий в этом процессе – важный этап на пути к разработке эффективных геропротекторов и преодолению

* Адресат для корреспонденции.

трудностей, которые старение приносит нашему обществу. В последние годы системный подход к изучению старения становится всё более популярным. Этому способствует, в частности, активное развитие методов высокопроизводительного секвенирования. Экспоненциальное накопление общедоступных омиксных данных приводит к необходимости их агрегации в специализированные базы данных, где они становятся более доступными для мета-анализа.

На сегодняшний день существует несколько баз данных, посвящённых геномике и транскриптомике старения, в том числе Aging Atlas [2], HAGR [3], Digital Ageing Atlas [4], AGEMAP [5] и Open Genes [6]. Некоторые из них предоставляют информацию не только о транскриптомике старения, но и о вмешательствах, увеличивающих продолжительность жизни, о фенотипе старения и эпигенетике. Однако при описании экспрессии генов эти базы, как правило, ограничиваются лишь списками соответствующих генов (например, дифференциально экспрессирующихся с возрастом). Насколько нам известно, на данный момент не существует баз данных, которые бы содержали количественную информацию, полученную на основе мета-анализа, об общих закономерностях изменения экспрессии генов с возрастом, включая видоспецифичные, тканеспецифичные и глобальные тенденции, и позволяли бы легко сравнивать и визуализировать эти данные для каждого гена на уровне отдельных наборов данных и отдельных образцов внутри каждого набора данных. Чтобы восполнить этот пробел, мы разработали базу данных AgeMeta, предоставляющую интерактивный пользовательский интерфейс для исследования возрастных биомаркеров генной экспрессии в конкретных тканях и видах млекопитающих, а также клеточных функций, обогащённых в транскриптомных паттернах старения. Наш ресурс представляет эту информацию как на уровне общих закономерностей, так и на уровне отдельных наборов данных. Для удобства использования мы разработали подробное руководство по каждому разделу базы данных и сделали её содержимое доступным для скачивания.

МАТЕРИАЛЫ И МЕТОДЫ

Полный список наборов данных, использованных для построения транскриптомных сигнатур старения млекопитающих и интегрированных в AgeMeta, включал следующие идентификаторы исследований из баз данных GEO [7], ArrayExpress [8] и SRA [9]: GSE9103, GSE123981, GSE3150, GSE6591, GSE74463, GSE53960, GSE66715, GSE11291, GSE34378, GSE27625, GSE12480, GSE36192, GSE1572, GSE28422,

GSE25941, GSE53890, GSE38718, GSE674, GSE17612, GSE21935, GSE362, GSE132040, E-MTAB-3374, PRJNA281127, PRJNA516151. Помимо этих исследований, мы также использовали данные GTEx [10]. Наборы данных содержали образцы RNA-Seq и микрочипов, полученные из различных тканей *Mus musculus*, *Rattus norvegicus* и *Homo sapiens*. Содержание тканей варьировалось у разных видов. 23 ткани, используемые для всех трёх видов, включали жировую ткань (в некоторых случаях обозначаемую отдельно как бурая, брыжеечная, гоная или подкожная жировые ткани), надпочечники, кровеносные сосуды, кость и костный мозг, головной мозг (в некоторых случаях обонятельные луковицы, мозжечок и лобная кора обозначались отдельно), пищевод, сердце, почки, печень, лёгкие, мышцы, нервы, поджелудочную железу, гипофиз, простату, слюнную железу, кожу, тонкий кишечник, селезёнку, семенник, щитовидную железу и цельную кровь (в некоторых случаях белые кровяные тельца обозначались отдельно). Аннотацию каждого набора данных, использованного в этом исследовании, включающую информацию о количестве образцов, возрастном диапазоне, линии, ткани и половой принадлежности, можно найти в дополнительных материалах (табл. S1 Приложения).

Процедура предварительной обработки данных для RNA-Seq и микрочипов подробно описана в работе Tyshkovskiy et al. [11]. Для обоих типов данных все идентификаторы генов стандартизировали в виде идентификаторов Entrez и, в случае образцов крыс и людей, переводили в идентификаторы мышиных ортологов, имеющих однозначное соответствие с исходными Entrez-идентификаторами. Эти два этапа предварительной обработки привели к тому, что некоторые гены были исключены из данных. Гены, которые не имели значений дифференциальной экспрессии более чем в 40% наборов данных, использованных для построения конкретной сигнатуры, также были отфильтрованы перед соответствующим мета-анализом. Таким образом, если определённый ген отсутствует в AgeMeta, это может быть вызвано его исключением на любом из вышеуказанных этапов обработки данных.

Логарифмы кратности изменения экспрессии генов (logFC), используемые для построения сигнатур старения, рассчитывали как коэффициент наклона в линейной модели, когда в выборке присутствовали образцы с более чем двумя уникальными возрастными, или как разность средних значений в противном случае, с использованием пакета limma для R [12]. Модель смешанных эффектов использовали для агрегирования изменений экспрессии из всех наборов данных в сигнатуры старения с помощью пакета metafor [13]. В качестве входных данных использовали logFC и соответствующие им

стандартные ошибки, а идентификатор исследования, ткань и вид вводили в модель как случайные эффекты. Анализ обогащения наборов генов проводили с использованием пакета fgsea [14] с 10 000 пермутаций для отдельных наборов данных и сигнатур старения. Для создания веб-сайта AgeMeta использовали пакет Shiny для R [15].

ОПИСАНИЕ БАЗЫ ДАННЫХ

Содержимое базы данных основано на семи транскриптомных сигнатурах старения, идентифицированных в нашем недавнем исследовании [11]. Транскриптомные сигнатуры представляют собой общие тенденции изменения экспрессии генов с возрастом, построенные путём мета-анализа 122 наборов данных из 26 исследований (таблица). Тканеспецифичные сигнатуры мозга, скелетных мышц и печени были получены из соответствующих наборов данных человека, мыши и крысы, а видоспецифичные сигнатуры обобщают данные по множеству различных тканей определённого вида. Глобальная сигнатура включает в себя все данные по трём видам и 23 тканям (см. раздел «Материалы и методы»).

База данных состоит из 5 основных модулей (рис. 1). Первый из них – это интерфейс для сравнения сигнатур, содержащий агрегированные количественные оценки возрастных изменений экспрессии (нормализованный logFC) в конкретной ткани и виде вместе с соответствующими показателями статистической значимости (*p*-value до и после поправки Бенджамини–Хохберга [16]) для каждого гена. Ресурс позволяет выбрать одну из сигнатур, а также сортировать и фильтровать гены в соответствии с их значимостью в выбранной сигнатуре. Кроме того, с помощью базы данных пользователь может сравнить значения дифференциальной экспрессии выбранных генов и их значимость между всеми 7 сигнатурами старения, используя правую часть интерактивной таблицы, где цветные стрелки рядом со значениями logFC иллюстрируют направление изменений экспрессии генов, а оттенки ячеек показывают статистическую значимость (рис. 2). Боковое меню

позволяет фильтровать гены по их присутствию и значимости во всех семи сигнатурах, а также на основании того, повышается или понижается их уровень экспрессии с возрастом в соответствии с каждой из сигнатур.

Если пользователь захочет проследить происхождение каждого значения logFC в интерактивной таблице, он может щёлкнуть по ячейке с соответствующим значением, что переведёт его на график модели смешанных эффектов, связанный с выбранным геном и сигнатурой (рис. 3, а). Модель смешанных эффектов использовалась для агрегирования данных дифференциальной экспрессии из отдельных наборов данных в сигнатуры (см. «Материалы и методы»). На графике каждая точка представляет один набор данных, то есть одну уникальную комбинацию вида, ткани, пола, линии и идентификатора исследования. Описание набора данных отображается при наведении курсора мыши на каждую точку. Красная линия показывает агрегированное среднее значение logFC для данного гена, представленное в таблице для сравнения сигнатур (это значение является результатом модели смешанных эффектов), а значение по оси ординат для каждой точки отображает logFC из соответствующего набора данных. Цвет и форму точек можно настроить для удобства различения видов, тканей, исследований, пола, платформ и типов моделей, используемых для получения значений дифференциальной экспрессии, что позволяет пользователю настроить график под свои потребности. Щелчок по точке на этом графике перенаправляет пользователя на график экспрессии для одного набора данных, который показывает, как экспрессия выбранного гена меняется с возрастом в выбранном наборе данных (рис. 3, б). Целью данного графика экспрессии является отображение необработанных данных, используемых для расчёта значения logFC и его стандартной ошибки для конкретного набора данных.

Последние два модуля базы данных в основном дублируют первые два, за исключением того, что вместо генов здесь представлены функциональные группы генов. Функциональное обогащение сигнатур и отдельных наборов данных проводили с помощью метода «Gene Set Enrichment Analysis» [17, 18],

Обзор семи транскриптомных сигнатур старения

Критерий	Человек	Мышь	Крыса	Мозг	Мышцы	Печень	Глобальная
Кол-во наборов данных	51	52	19	25	17	11	122
Кол-во генов	15 959	13 459	10 382	16 125	15 339	12 805	15 170
Кол-во генов, экспрессия которых значимо ассоциирована с возрастом (<i>p</i> adjusted < 0,05)	649	184	501	1526	379	1219	7

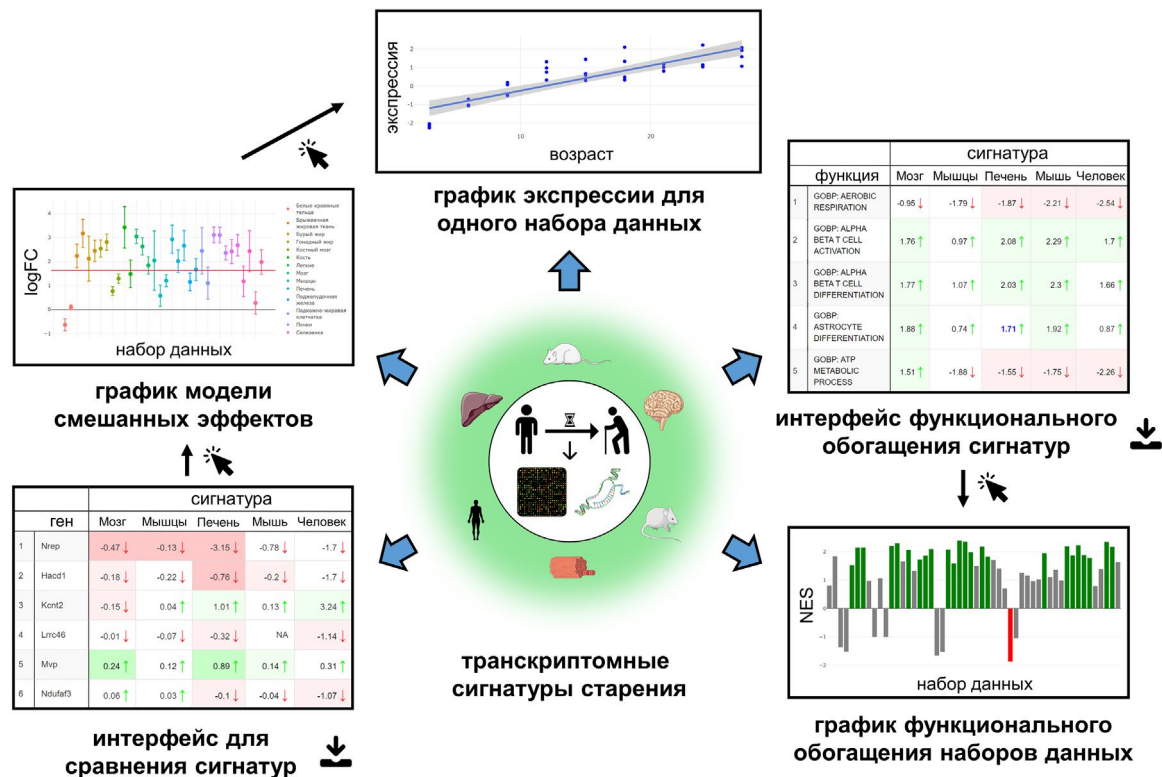


Рис. 1. Структура базы данных AgeMeta. AgeMeta представляет собой визуальный интерфейс, посвящённый 7 транскриптомным сигнатурам старения. Он содержит две интерактивные таблицы, отображающие количественные возрастные изменения в транскрипте для индивидуальных генов (интерфейс для сравнения сигнатур) и функциональных групп генов (интерфейс функционального обогащения сигнатур), а также позволяет сравнивать между собой сигнатуры. AgeMeta также содержит три типа графиков, на которых представлены: дифференциальная экспрессия гена в сигнатуре на уровне различных наборов данных (график модели смешанных эффектов), изменение экспрессии гена с возрастом в отдельном наборе данных (график экспрессии для одного набора данных) и показатели обогащения функциональных групп генов в отдельных наборах данных, связанных с определённой сигнатурой (график функционального обогащения наборов данных). Щелчок по ячейкам таблиц переводит пользователя к соответствующим графикам, а щелчок по точкам на графике модели смешанных эффектов переводит пользователя к графику экспрессии, связанному с выбранным набором данных. Обе таблицы доступны для скачивания в отдельном разделе базы данных



Рис. 2. Интерфейс для сравнения сигнатур. Цвета стрелок отображают направление изменений экспрессии генов в каждой сигнатуре (названия сигнатур указаны в заголовках столбцов), а цвета ячеек передают статистическую значимость этих изменений. Если значение p-value после поправки на множественное тестирование оказывается ниже мягкого порога, ячейка окрашивается тусклым оттенком, а если ниже жёсткого порога – более ярким оттенком. Оба порога могут быть настроены в меню в левой части страницы. Все ячейки в правой части таблицы со значениями logFC доступны для нажатия. Нажатие на них перенаправляет пользователя на соответствующий график модели смешанных эффектов

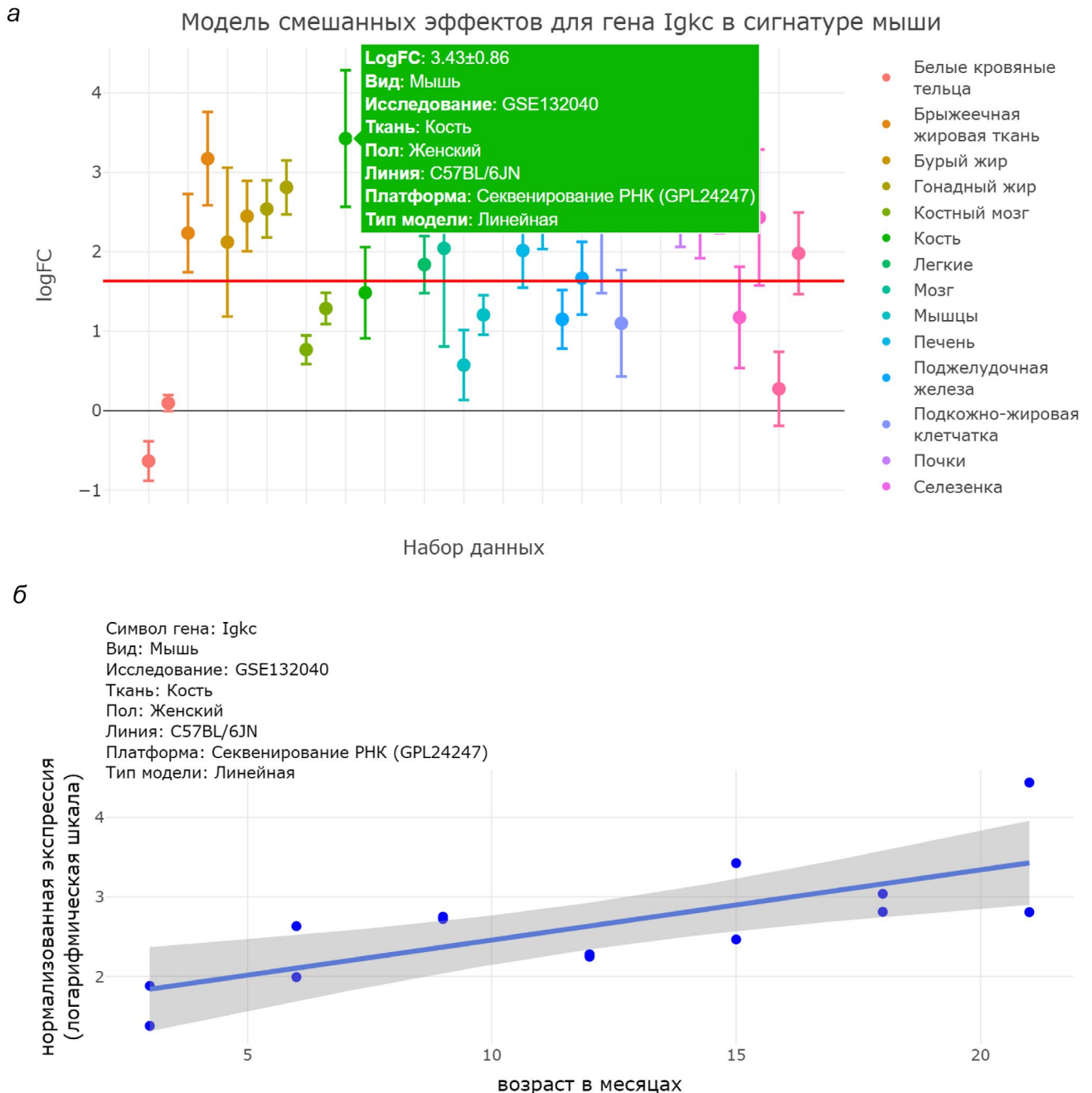


Рис. 3. Примеры графиков модели смешанных эффектов и экспрессии. *а* – Пример графика модели смешанных эффектов. Каждая точка представляет собой один набор данных. Форма и цвет точек обозначают вид и ткань соответственно. Данные отображают средний нормализованный наклон/logFC $\pm$ SE. При наведении курсора на точку в текстовом поле появляется подробная информация о соответствующем наборе данных. *б* – Пример графика экспрессии для одного набора данных, соответствующего точке, выбранной на графике модели смешанных эффектов. Наклон линии регрессии соответствует значению logFC для гена в выбранном наборе данных (координата по оси ординат выбранной точки), а ширина затенённой области вокруг линии отражает стандартную ошибку наклона прямой

в результате чего были рассчитаны количественные нормализованные показатели обогащения (NES) и *p*-value после поправки Бенджамини–Хохберга. В модуле интерфейса функционального обогащения сигнатур пользователь может сравнить показатели обогащения и соответствующие статистические значимости для всех семи сигнатур.

Функциональные группы генов можно фильтровать по их присутствию, статистической значимости, направлению изменения экспрессии в соответствии с сигнатурами, а также на основе базы данных, к которой относится эти группы. Используемые базы включали в себя GO Biological Process [19, 20], KEGG [21–23], Reactome [24] и Biocarta [25].

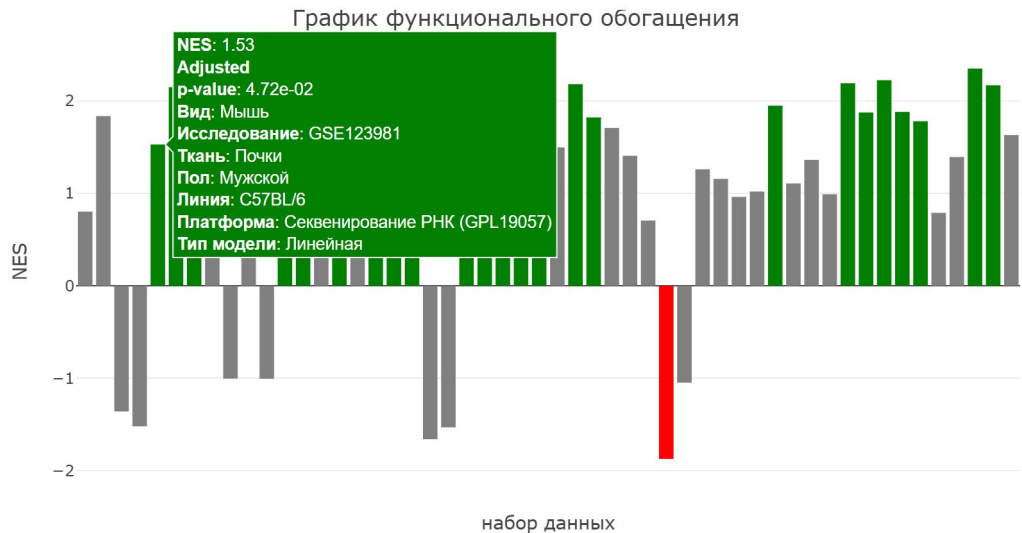


Рис. 4. Пример графика функционального обогащения наборов данных. График отображает функцию «GOBP: обработка и презентация антигенов» в сигнатуре мыши. Столбцы окрашены в зависимости от статистической значимости: зелёный и красный цвета указывают на статистически значимое повышение и понижение экспрессии функциональной группы генов с возрастом соответственно, а серые столбцы обозначают незначимое обогащение. Подробное описание каждого набора данных отображается в текстовом поле при наведении курсора мыши на соответствующий столбец графика

Поскольку многие функциональные группы генов вложены друг в друга, для упрощения навигации мы добавили функцию, позволяющую одновременно отобразить в таблице все родительские и дочерние наборы генов для выбранного функционального термина, т.е. группы, которые содержат выбранный набор генов или являются подмножествами этого набора соответственно. При нажатии на название функционального термина таблица автоматически фильтруется, в результате чего остаётся только выбранный термин, а также его родительские и дочерние термины. Кнопка «Вернуться к полной таблице» восстанавливает таблицу в исходное состояние со всеми функциональными группами из базы данных.

Щелчок по ячейке таблицы направляет пользователя к графику функционального обогащения наборов данных, на котором отображаются оценки обогащения и статистическая значимость для всех отдельных наборов данных, соответствующих выбранному функциональному термину и сигнатуре (рис. 4). Хотя агрегированные показатели функционального обогащения сигнатур не были получены из оценок отдельных наборов данных, этот график даёт общее представление о согласованности возрастного повышения или понижения экспрессии генов, связанных с выбранным путём, в разных наборах данных и помогает обосновывать значения в интерактивной таблице. Каждый столбец на графике представляет один набор данных, описание которого отображается при наведении курсора мыши. Зелёные и красные столбцы обозначают статистически значимое

повышение или понижение экспрессии генов с возрастом соответственно, а серые столбцы отражают показатели обогащения, не достигшие порога статистической значимости, установленного в боковом меню (по умолчанию порог на p -value после поправки на множественное тестирование равен 0,1).

Для удобства пользователей мы разработали раздел «Руководство», где подробно описан каждый модуль базы данных с примерами таблиц и графиков. Кроме того, ресурс доступен на двух языках: английском и русском. Чтобы упростить независимый анализ представленной в базе информации, мы добавили функцию скачивания данных дифференциальной экспрессии и данных функционального обогащения. Доступ к ним можно получить через раздел «Скачать данные».

ОБСУЖДЕНИЕ

В базе данных AgeMeta представлена информация об изменении экспрессии нескольких тысяч генов, связанных со старением. С помощью инструментов, встроенных в AgeMeta, можно легко визуализировать паттерны дифференциальной экспрессии, а также их межтканевые или межтканевые различия для любого из этих генов или для их функциональных групп. Интерактивные таблицы и графики позволяют анализировать информацию на уровне транскриптомных сигналов и отдельных наборов данных. Успешное применение подобных сигнатур в недавних работах

для предсказания и описания новых интервенций, направленных на увеличение продолжительности жизни и омоложение [11, 26], свидетельствует о том, что сигнатуры старения, включённые в AgeMeta, могут способствовать открытию и разработке новых потенциальных геропротекторов.

На сегодняшний день существует несколько баз данных по геномике и транскриптомике старения, поэтому стоит выделить наиболее важные различия между базой AgeMeta и наиболее схожими с ней ресурсами. GenAge, являющаяся частью собрания баз данных, известного как «Human Ageing Genomic Resources» или HAGR [3], фокусируется на описании воздействия геномных интервенций (например, нокаутов генов) на продолжительность жизни таких модельных организмов, как *Mus musculus*. Однако связь этих эффектов с возрастными транскриптомными изменениями нетривиальна: геномные интервенции, приводящие к увеличению продолжительности жизни, не всегда оказывают влияние, противоположное изменениям экспрессии генов при старении. Отчасти это можно объяснить тем, что некоторые возрастные изменения экспрессии генов являются вредными для организма, а другие – адаптивными, оказывающими положительное влияние на здоровье стареющих животных [11]. Примером такого эффекта является экспрессия гена *Pparg*, которая значительно повышается с возрастом в тканях мышей, в то время как нокаут этого гена приводит к существенному снижению продолжительности их жизни [27].

Существуют и другие базы данных, посвящённые транскриптомике старения, такие как Aging Atlas [2] и AGEMAP [5]. Обе они содержат информацию о дифференциальной экспрессии генов из отдельных исследований (AGEMAP включает только одно исследование). Однако ни одна из них не предоставляет значений дифференциальной экспрессии, полученных на основе мета-анализа, агрегированных по нескольким исследованиям с различными тканями и видами животных, что является ключевой особенностью AgeMeta. Aging Atlas содержит также мультиомиксные данные, такие как белок-белковые взаимодействия, данные ChIP-Seq и транскриптомики одиночных клеток, описание которых выходит за рамки нашего исследования. Другая база данных, Open Genes [6], содержит описания генов с различными типами

ассоциаций со старением, включая нокауты, влияющие на продолжительность жизни, и возрастную дифференциальную экспрессию. Однако эта база данных также не содержит результатов количественного мета-анализа.

Следует отметить несколько ограничений нашей работы. Во-первых, сигнатуры старения, используемые в AgeMeta, были основаны на линейных моделях дифференциальной экспрессии. Это может означать, что некоторые важные нелинейные возрастные изменения экспрессии генов могли оказаться не идентифицированы как статистически значимые в данной работе. Во-вторых, наше исследование ограничивалось анализом изменений экспрессии индивидуальных генов, хотя в других работах была отмечена дифференциальная ко-экспрессия, наблюдаемая с возрастом у различных видов млекопитающих [28–30]. Интеграция мета-анализа нелинейной дифференциальной экспрессии и паттернов ко-экспрессии генов в нашу базу данных может в будущем дополнительно увеличить её полноту и применимость, а также способствовать более глубокому пониманию научным сообществом молекулярных механизмов старения млекопитающих.

Дополнительные материалы. Приложение к статье опубликовано на сайте журнала «Биохимия» (<https://biochemistry.moscow.com>).

Вклад авторов. А.Э.Т., С.Е.Д. и М.А.Б. разработали план исследования; С.А.Т. и А.Э.Т. проектировали базу данных; С.А.Т. выполнил анализ данных и написал код для базы данных; С.А.Т., М.А.Б., В.Н.Г., С.Е.Д. и А.Э.Т. интерпретировали данные; А.Э.Т. руководил исследованием; С.А.Т. и А.Э.Т. написали рукопись при участии всех остальных авторов. Все авторы прочли и одобрили окончательную версию рукописи.

Финансирование. Работа выполнена при поддержке Российского научного фонда (грант № 23-14-00218, С.Е.Д. – обработка данных и создание транскриптомных сигнатур) и Open Longevity (С.А.Т. и А.Э.Т. – разработка базы данных).

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

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Harman, D. (1991) The aging process: Major risk factor for disease and death, *Proc. Natl. Acad. Sci. USA*, **88**, 5360-5363, <https://doi.org/10.1073/pnas.88.12.5360>.
2. Consortium, A. A. (2021) Aging Atlas: a multi-omics database for aging biology, *Nucleic Acids Res.*, **49**, 825-830, <https://doi.org/10.1093/nar/gkaa894>.

3. Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., Diana, E., Lehmann, G., Toren, D., Wang, J., De Magalh, P., and Fraielfeld, E. (2018) Human Ageing Genomic Resources: new and updated databases, *Nucleic Acids Res.*, **46**, 1083-1090, <https://doi.org/10.1093/nar/gkx1042>.
4. Craig, T., Smelick, C., Tacutu, R., Wuttke, D., Wood, S. H., Stanley, H., Janssens, G., Savitskaya, E., Moskalev, A., De Magalh, P., and Arking, R. (2015) The Digital Ageing Atlas: integrating the diversity of age-related changes into a unified resource, *Nucleic Acids Res.*, **43**, 873-878, <https://doi.org/10.1093/nar/gku843>.
5. Zahn, J. M., Poosala, S., Owen, A. B., Ingram, D. K., Lustig, A., Carter, A., Weeraratna, A. T., Taub, D. D., Gorospe, M., Mazan-Mamczarz, K., Lakatta, E. G., Boheler, K. R., Xu, X., Mattson, M. P., Falco, G., Ko, M. S. H., Schlessinger, D., Firman, J., Kummerfeld, S. K., Wood, H. W., 3rd, Zonderman, A. B., Kim, S. K., and Becker, K. G. (2007) AGEMAP: a gene expression database for aging in mice, *PLoS Genet.*, **3**, e201, <https://doi.org/10.1371/journal.pgen.0030201>.
6. Rafikova, E., Nemirovich-Danchenko, N., Ogmen, A., Parfenenkov, A., Velikanova, A., Tikhonov, S., Peshkin, L., Rafikov, K., Spiridonova, O., Belova, Y., Glinin, T., Egorova, A., and Batin, M. (2023) Open Genes – a new comprehensive database of human genes associated with aging and longevity, *Nucleic Acids Res.*, **52**, D950-D962, <https://doi.org/10.1093/nar/gkad712>.
7. Edgar, R., Domrachev, M., and Lash, A. E. (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository, *Nucleic Acids Res.*, **30**, 207-210, <https://doi.org/10.1093/nar/30.1.207>.
8. Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, N., Coulson, R., Farne, A., Holloway, E., Kolesnykov, N., Lilja, P., Lukk, M., Mani, R., Rayner, T., Sharma, A., William, E., Sarkans, U., and Brazma, A. (2007) Array-Express – a public database of microarray experiments and gene expression profiles, *Nucleic Acids Res.*, **35**, 747-50, <https://doi.org/10.1093/nar/gkl995>.
9. Leinonen, R., Sugawara, H., and Shumway, M. (2011) The sequence read archive, *Nucleic Acids Res.*, **39**, 2010-2012, <https://doi.org/10.1093/nar/gkq1019>.
10. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J., Fleming, J., Siminoff, L., Traino, H., Mosavel, M., Barker, L., Jewell, S., Rohrer, D., Maxim, D., Filkins, D., Harbach, P., Cortadillo, E., Berghuis, B., Turner, L., Hudson, E., Feenstra, K., Sobin, L., Robb, J., Branton, P., Korzeniewski, G., Shive, C., Tabor, D., Qi, L., Groch, K., Nampally, S., Buia, S., Zimmerman, A., Smith, A., Burges, R., Robinson, K., Valentino, K., Bradbury, D., Cosentino, M., Diaz-Mayoral, N., Kennedy, M., Engel, T., Williams, P., Erickson, K., Ardlie, K., Winckler, W., Getz, G., DeLuca, D., MacArthur, D., Kellis, M., Thomson, A., Young, T., Gelfand, E., Donovan, M., Meng, Y., Grant, G., Mash, D., Marcus, Y., Basile, M., Liu, J., Zhu, J., Tu, Z., Cox, N. J., Nicolae, D. L., Gamazon, E. R., Im, H. K., Konkashbaev, A., Pritchard, J., Stevens, M., Flutre, T., Wen, X., Dermitzakis, E. T., Lappalainen, T., Guigo, R., Monlong, J., Sammeth, M., Koller, D., Battle, A., Mostafavi, S., McCarthy, M., Rivas, M., Maller, J., Rusyn, I., Nobel, A., Wright, F., Shabalin, A., Feolo, M., Sharopova, N., Sturcke, A., Paschal, J., Anderson, J. M., Wilder, E. L., Derr, L. K., Green, E. D., Struewing, J. P., Temple, G., Volpi, S., Boyer, J. T., Thomson, E. J., Guyer, M. S., Ng, C., Abdallah, A., Colantuoni, D., Insel, T. R., Koester, S. E., Little, A. R., Bender, P. K., Lehner, T., Yao, Y., Compton, C. C., Vaught, J. B., Sawyer, S., Lockhart, N. C., Demchok, J., and Moore, H. F. (2013) The Genotype-Tissue Expression (GTEx) project, *Nat. Genet.*, **45**, 580-585, <https://doi.org/10.1038/ng.2653>.
11. Tyshkovskiy, A., Ma, S., Shindyapina, A. V., Tikhonov, S., Lee, S. G., Bozaykut, P., Castro, J. P., Seluanov, A., Schork, N. J., Gorbunova, V., Dmitriev, S. E., Miller, R. A., and Gladyshev, V. N. (2023) Distinct longevity mechanisms across and within species and their association with aging, *Cell*, **186**, 2929-2949.e20, <https://doi.org/10.1016/j.cell.2023.05.002>.
12. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Res.*, **43**, e47, <https://doi.org/10.1093/nar/gkv007>.
13. Viechtbauer, W. (2010) Conducting meta-analyses in R with the metafor, *J. Stat. Softw.*, **36**, 1-48, <https://doi.org/10.18637/jss.v036.i03>.
14. Korotkevich, G., Sukhov, V., and Sergushichev, A. (2021) Fast gene set enrichment analysis, *bioRxiv*, <https://doi.org/10.1101/060012>.
15. Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., and Borges, B. (2022) Shiny: Web Application Framework for R.
16. Benjamini, Y., Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Ser. B*, **57**, 289-300, <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
17. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. USA*, **102**, 15545-15550, <https://doi.org/10.1073/pnas.0506580102>.
18. Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegel-

- man, B., Lander, E. S., Hirschhorn, J. N., Altshuler, D., and Groop, L. C. (2003) PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes, *Nat. Genet.*, **34**, 267-273, <https://doi.org/10.1038/ng1180>.
19. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene Ontology: tool for the unification of biology, *Nat. Genet.*, **25**, 25-29, <https://doi.org/10.1038/75556>.
20. Carbon, S., Douglass, E., Good, B. M., Unni, D. R., Harris, N. L., Mungall, C. J., Basu, S., Chisholm, R. L., Dodson, R. J., Hartline, E., Fey, P., Thomas, P. D., Albou, L. P., Ebert, D., Kesling, M. J., Mi, H., Muruganujan, A., Huang, X., Mushayahama, T., LaBonte, S. A., Siegele, D. A., Antonazzo, G., Attrill, H., Brown, N. H., Garapati, P., Marygold, S. J., Trovisco, V., dos Santos, G., Falls, K., Tabone, C., Zhou, P., Goodman, J. L., Strelets, V. B., Thurmond, J., Garmiri, P., Ishtiaq, R., Rodríguez-López, M., Acencio, M. L., Kuiper, M., Lægreid, A., Logie, C., Lovering, R. C., Kramarz, B., Sav-erimuttu, S. C. C., Pinheiro, S. M., Gunn, H., Su, R., Thurlow, K. E., Chibucos, M., Giglio, M., Nadendla, S., Munro, J., Jackson, R., Duesbury, M. J., Del-Toro, N., Meldal, B. H. M., Paneerselvam, K., Perfetto, L., Porras, P., Orchard, S., Shrivastava, A., Chang, H. Y., Finn, R. D., Mitchell, A. L., Rawlings, N. D., Richardson, L., Sangrador-Vegas, A., Blake, J. A., Christie, K. R., Dolan, M. E., Drabkin, H. J., Hill, D. P., Ni, L., Sitnikov, D. M., Harris, M. A., Oliver, S. G., Rutherford, K., Wood, V., Hayles, J., Bähler, J., Bolton, E. R., de Pons, J. L., Dwinell, M. R., Hayman, G. T., Kaldunski, M. L., Kwitek, A. E., Lauderkind, S. J. F., Plasterer, C., Tutaj, M. A., VEDI, M., Wang, S. J., D'Eustachio, P., Matthews, L., Balhoff, J. P., Aleksander, S. A., Alexander, M. J., Cherry, J. M., Engel, S. R., Gondwe, F., Karra, K., Miyasato, S. R., Nash, R. S., Simison, M., Skrzypek, M. S., Weng, S., Wong, E. D., Feuermann, M., Gaudet, P., Morgat, A., Bakker, E., Berardini, T. Z., Reiser, L., Subramaniam, S., Huala, E., Arighi, C. N., Auchincloss, A., Axelsen, K., Argoud-Puy, G., Bateman, A., Blatter, M. C., Boutet, E., Bowler, E., Breuza, L., Bridge, A., Britto, R., Bye-A-Jee, H., Casas, C. C., Coudert, E., Denny, P., Es-Treicher, A., Famiglietti, M. L., Georghiou, G., Gos, A. N., Gruaz-Gumowski, N., Hatton-Ellis, E., Hulo, C., Ignatchenko, A., Jungo, F., Laiho, K., Le Mercier, P., Lieberherr, D., Lock, A., Lussi, Y., MacDougall, A., Ma-Grane, M., Martin, M. J., Masson, P., Natale, D. A., Hyka-Nouspikel, N., Orchard, S., Pedruzzi, I., Pourcel, L., Poux, S., Pundir, S., Rivoire, C., Speretta, E., Sundaram, S., Tyagi, N., Warner, K., Zaru, R., Wu, C. H., Diehl, A. D., Chan, J. N., Grove, C., Lee, R. Y. N., Muller, H. M., Raciti, D., van Auken, K., Sternberg, P. W., Berriman, M., Paulini, M., Howe, K., Gao, S., Wright, A., Stein, L., Howe, D. G., Toro, S., Westerfield, M., Jaiswal, P., Cooper, L., and Elser, J. (2021) The gene ontology resource: enriching a Gold mine, *Nucleic Acids Res.*, **49**, D325-334, <https://doi.org/10.1093/nar/gkaa1113>.
21. Kanehisa, M. (2000) KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res.*, **28**, 27-30, <https://doi.org/10.1093/nar/28.1.27>.
22. Kanehisa, M. (2019) Toward understanding the origin and evolution of cellular organisms, *Protein Sci.*, **28**, 1947-1951, <https://doi.org/10.1002/pro.3715>.
23. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021) KEGG: Integrating viruses and cellular organisms, *Nucleic Acids Res.*, **49**, D545-551, <https://doi.org/10.1093/nar/gkaa970>.
24. Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Weiser, J., Wu, G., Stein, L., Hermjakob, H., and D'Eustachio, P. (2020) The reactome pathway knowledgebase, *Nucleic Acids Res.*, **48**, D498-D503, <https://doi.org/10.1093/nar/gkz1031>.
25. Rouillard, A. D., Gundersen, G. W., Fernandez, N. F., Wang, Z., Monteiro, C. D., McDermott, M. G., and Ma'ayan, A. (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins, *Database (Oxford)*, **2016**, 1-16, <https://doi.org/10.1093/database/baw100>.
26. Zhang, B., Lee, D. E., Trapp, A., Tyshkovskiy, A., Lu, A. T., Bareja, A., Kerepesi, C., McKay, L. K., Shindyapina, A. V., Dmitriev, S. E., Baht, G. S., Horvath, S., Gladyshev, V. N., and White, J. P. (2023) Multi-omic rejuvenation and lifespan extension on exposure to youthful circulation, *Nat. Aging*, **3**, 948-964, <https://doi.org/10.1038/s43587-023-00451-9>.
27. Argmann, C., Dobrin, R., Heikkinen, S., Auburtin, A., Pouilly, L., Cock, T. A., Koutnikova, H., Zhu, J., Schadt, E. E., and Auwerx, J. (2009) Ppar γ 2 is a key driver of longevity in the mouse, *PLoS Genet.*, **5**, e1000752, <https://doi.org/10.1371/journal.pgen.1000752>.
28. Bernard, G., Teulière, J., Lopez, P., Corel, E., Lapointe, F. J., and Baptiste, E. (2022) Aging at evolutionary crossroads: longitudinal gene co-expression network analyses of proximal and ultimate causes of aging in bats, *Mol. Biol. Evol.*, **39**, msab302, <https://doi.org/10.1093/molbev/msab302>.
29. Southworth, L. K., Owen, A. B., and Kim, S. K. (2009) Aging mice show a decreasing correlation of gene expression within genetic modules, *PLoS Genet.*, **5**, e1000776, <https://doi.org/10.1371/journal.pgen.1000776>.
30. Hu, Y., Xu, Y., Mao, L., Lei, W., Xiang, J., Gao, L., Jiang, J., Huang, L., Luo, O. J., Duan, J., and Chen, G. (2021) Gene expression analysis reveals age and ethnicity signatures between young and old adults in human PBMC, *Front. Aging*, **2**, 1-16, <https://doi.org/10.3389/fragi.2021.797040>.

AgeMeta: QUANTITATIVE GENE EXPRESSION DATABASE OF MAMMALIAN AGING

S. Tikhonov^{1,2}, M. Batin³, V. N. Gladyshev⁴, S. E. Dmitriev^{1,2}, and A. Tyshkovskiy^{1,4*}

¹ *Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, 119992 Moscow, Russia*

² *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, 119234 Moscow, Russia*

³ *Open Longevity, 15260 Ventura Blvd, STE 2230, Sherman Oaks, CA 91403, USA*

⁴ *Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA; e-mail: atyshkovskii@bwh.harvard.edu*

AgeMeta is a database that provides systemic and quantitative description of mammalian aging at the level of gene expression. It encompasses transcriptomic changes with age across various tissues of humans, mice, and rats, based on a comprehensive meta-analysis of 122 publicly available gene expression datasets from 26 studies. AgeMeta provides an intuitive visual interface for quantification of aging-associated transcriptomics at the level of individual genes and functional groups of genes, allowing easy comparison among various species and tissues. Additionally, all the data in the database can be downloaded and analyzed independently. Overall, this work contributes to the understanding of the complex network of biological processes underlying mammalian aging and supports future advancements in this field. AgeMeta is freely available at: <https://age-meta.com/>.

Keywords: aging, ageing, gene expression, differential expression, transcriptomics, mammalian, meta-analysis, AgeMeta, age-meta, database, GSEA