

МОЛЕКУЛЯРНАЯ БИОЛОГИЯ

УДК 556.3;540.4

DOI: 10.54503/0321-1339-2024.124.3-4-42

А.С. Юдина

Интеграция NGS данных комплексного геномного профилирования солидной и жидкостной биопсии для улучшения молекулярной диагностики онкологических заболеваний

(Представлено чл.-кор. НАН РА Р. М. Арутюняном 06/IX 2024)

Ключевые слова: секвенирование нового поколения, прецизионная онкология, выявление мутаций, цодНК.

Введение. В современной практике терапии онкологических заболеваний возрастающую роль играет терапия, направленная на раковые клетки – таргетированная терапия [1, 2], которая может использоваться как основное лечение или как компонент к стандартным методам: химиотерапии, хирургическому удалению опухоли, лучевой терапии. Американское сообщество по изучению онкологических заболеваний выделяет следующие направления таргетированной терапии: ингибиторы ангиогенеза, моноклональные антитела к специфическим поверхностным биомаркерам раковых клеток, протеасомные ингибиторы и разобщители проведения внутриклеточного сигналинга [3, 4]. Перечисленные таргетированные подходы основываются на детальном изучении биомаркеров опухоли: альтерациях в ДНК, во внутриклеточных и поверхностных белках, в связи с этим методы молекулярной диагностики должны соответствовать требованиям области.

В случае солидной опухоли, один из основных способов поиска биомаркеров – анализ клеток, взятых напрямую из опухоли (тканевая биопсия). Секвенирование нового поколения (NGS) на полногеномных или таргетных генных панелях позволяет выявлять мутации в драйверных генах (*KRAS*, *BRAF*, *PIK3CA*, *ALK* и т.д.), для которых существуют терапии или проводятся клинические испытания [5, 6], а также рассчитывать комплексные метрики, такие как опухолевая мутационная нагрузка или оценка микросателлит-

ной нестабильности, для назначения ингибиторов иммунных контрольных точек [7, 8].

Для злокачественных заболеваний, к которым затруднен доступ для взятия биопсии, или задачи продолжительного мониторинга прогресса опухоли в момент или после лечения используется процедура анализа жидкостной биопсии (в большинстве случаев образца крови). По данным NGS внеклеточной циркулирующей опухолевой ДНК (цОДНК) возможно рассчитывать спектр биомаркеров, аналогичных солидной биопсии – геномные точечные альтерации, структурные варианты, мутационная нагрузка и т.д., а также специфичные для данного материала биомаркеры – паттерн метилирований CpG островков, распределение длин фрагментов ДНК, вирусные последовательности ДНК, уровень опухолевой ДНК в общей фракции внеклеточной ДНК [9, 10].

Проведенные исследования при сравнении результатов тестовых платформ для солидных и жидкостных биопсий сообщают о низком пересечении биомаркеров – 8,6% [11], 7,7% [12], 20% [13] и более высоких показателях пересечения при стратификации случаев по диагнозам и стадиям заболевания 77,6% [14] и 36 – 76% [15].

В настоящем исследовании была произведена оценка сходимости данных поиска биомаркеров на основе точечных мутаций и коротких делеций и вставок (инделов) по данным NGS тканевой и жидкостной биопсии с дополнительной проверкой мутаций в транскриптом опухоли с целью увеличения точности молекулярной диагностики для улучшения клинической интерпретации находок.

Результаты. Перед началом анализа в качестве базовой характеристики разных платформ для поиска биомаркеров была произведена оценка количества артефактов, которые необходимо исключать из данных перед их клинической интерпретацией. Для использованной когорты пациентов в данных NGS солидной биопсии истинные мутации составляют 38,6% (ДИ 95 – 33,6%–43,7%), для жидкостной биопсии – 1,1% (ДИ 95 – 0,9%–1,3%).

После фильтрации данных от артефактов прямое пересечение клинически значимых мутаций составило 27,1% (ДИ 95 – 23,2%–30,0%) для NGS ДНК тканевой и жидкостной биопсий и 82,2% (ДИ 95 – 79,4%–85,1%) для NGS ДНК и РНК тканевой биопсии (рис. А). Распределение типов мутаций на общем пересечении ДНК ткани с ДНК плазмы или РНК ткани выглядит схожим образом (рис. Б) с увеличенной долей (>75% суммарно) мутаций, приводящих к заменам аминокислот или не изменяющих белок мутаций. Ожидаемо, доля мутации интронов и регионов сплайсинга снижена для пересечения с РНК, относительно ДНК плазмы, при этом мутации, влияющие на длину транскрипта (мутации в стоп-кодонах), в равной доле присутствуют в обоих типах данных – 9,5% в пересечении с ДНК плазмы, 11% в пересечении с РНК ткани (рис. Б). Корреляционный анализ аллельных частот (рис. В, Г) выявил слабую зависимость между тканевой и жидкостной

биопсиями ($R: 0,41$, $p\text{-value} < 0,001$) и более уверенную корреляцию между РНК и ДНК ($R: 0,59$, $p\text{-value} < 0,001$) тканевой биопсии.

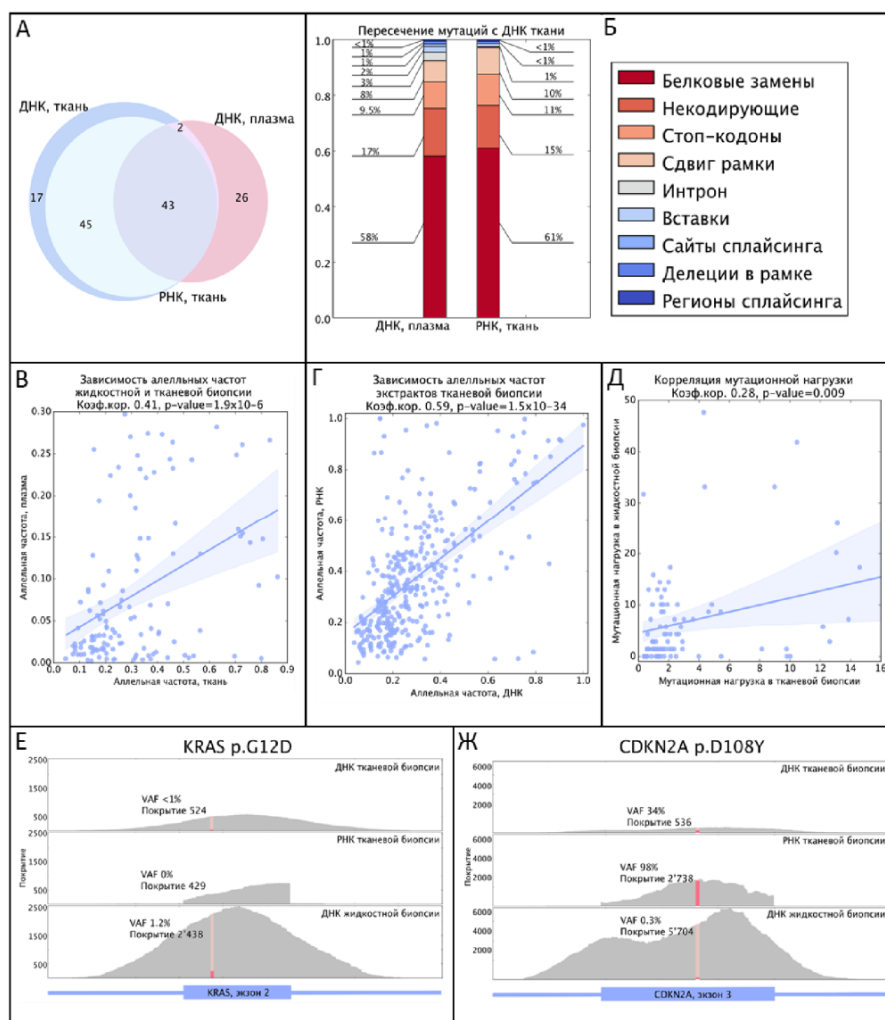


Рис. А – пересечение мутаций, задетектированных на разных платформах; Б – распределение типов мутаций в пересечении ДНК ткани с ДНК плазмы и РНК ткани; В – зависимость аллельных частот мутаций горячих точек в пересечении тканевой и жидкостной биопсии; Г – зависимость аллельных частот мутаций горячих точек в пересечении ДНК и РНК тканевой биопсии; Д – корреляция мутационной нагрузки (мут. / Mb) между тканевой и жидкостной биопсией; Е, Ж – покрытие и аллельная частота мутации KRAS p.G12D и CDKN2A p.D108Y в данных ДНК жидкостной биопсии и ДНК и РНК тканевой биопсии.

Для оценки сходимости по общим находкам был проведен анализ мутационной нагрузки, рассчитанной по солидной и жидкостной биопсиям. Коэффициент корреляции составил 0,28 ($p\text{-value} = 0,009$) для 87 пациентов

без учета выбросов (рис. Д). Оценка специфичности и отрицательная прогностическая ценность (TNR) для генов, вошедших в панели для обеих платформ, составили 96,5% (ДИ 95 – 95,9%–97,2%) и 98,1% (ДИ 95 – 97,7%–98,5%).

При объединении данных от разных платформ – тканевая и жидкостная биопсия, а также анализ экзона и транскриптома – количество пациентов, для которых были найдены клинически значимые мутации, возросло с 51 при стандартном анализе NGS ДНК тканевой биопсии до 58. Увеличение количества мутаций для клинического анализа происходило за счет подтверждения найденных точечных замен и инделов в данных цоДНК и/или NGS РНК (рис. Е, Ж). Примечательно, интеграция данных позволяет делать выводы о гетерогенности опухолевого состава – так, например, мутация *KRAS* p.G12D (рис. Е) детектируется в данных жидкостной биопсии, в то время как в образце ткани находится на уровне шума. Мутация *CDKN2A* p.D108Y (рис. Ж) детектируется во всех типах данных, но в транскриптоме ее аллельная частота повышена более чем в 2 раза относительно ДНК. Увеличенная экспрессия альтернативного аллеля косвенно свидетельствует об изменении числа копий гена.

Обсуждение. Применение NGS для клинической интерпретации влечет за собой повышенные требования к их контролю качества. Именно поэтому изучению артефактов и способам их фильтрации уделяется особое внимание. Артефакты в данных NGS имеют разную природу. Так, при анализе экстрактов тканевых биопсий из парафиновых блоков, важно учитывать возможность присутствия артефактов фиксации формалином, которые по геномному контексту схожи со спонтанным дезаминированием метилцитозина [16]. Кроме этого, отличительной чертой подобных артефактов является низкая аллельная частота (до 5%), что нужно принимать во внимание при поиске биомаркеров по NGS данным тканевых биопсий [17]. Для NGS по жидкостной биопсии основным источником артефактов являются ошибки ПЦР, возникающие на аллельной частоте, схожей с истинными мутациями (до 1%), и имеющие смещенное распределение по прямым и обратным прочтениям, что корректируется дополнительным шагом биоинформатического анализа – дедупликацией по уникальным молекулярным идентификаторам [18]. Подтверждение наблюдений о повышенной вероятности артефактов видно в данных использованной когорты – уровень ложноположительных находок в сырых данных цоДНК 98%, из-за чего при объединении находок разных платформ эти данные играют скорее второстепенную роль и находки в них верифицируются по данным тканевой биопсии. Данные РНК секвенирования в свою очередь подвержены ошибкам на концах прочтений, возникающим на этапе выравнивания из-за особенностей обработки экзонных последовательностей [19]. Именно поэтому поиск мутаций по РНК чаще всего в исследованиях используется как компонент к стандартным методам для подтверждения экспрессии вариантов.

Наблюдаемые результаты по пересечению биомаркерных мутаций по данным тканевой и жидкостной биопсии согласуются с опубликованными исследованиями [11–15]. В настоящем исследовании не производился отбор пациентов по стадии заболевания, однако распределение в основном было смещено к II–III стадиям, именно поэтому наблюдаемое пересечение находок на 27% выше среднего ожидаемого. Пересечение РНК и ДНК ниже 90% даже при условии общности биопсии, что может объясняться аллель специфической экспрессией генов [20]. Корреляция уровня мутационной нагрузки между кровью и тканью выражена слабо ($R: 0.28$), что также воспроизводит результаты и более масштабных исследований ($n=1137$) [21]. Однако исследования на диагнозе немелкоклеточного рака легкого показывают, что оценка мутационной нагрузки индивидуально по цоДНК может быть самостоятельным биомаркером [22, 23].

Кроме чувствительности на низких лимитах детекции для тестовых платформ также важна специфичность, т.к. при назначении таргетной терапии с учетом результатов геномного профилирования много внимания уделяется определению статуса мутированности гена. Так, например, при блокаде иммунных контрольных точек (PD-L1) для пациентов с немелкоклеточным раком легкого важно принимать во внимание, чтобы гены *EGFR* и *ALK*, продукты которых регулируют рост и деление клеток, не содержали мутации [24]. Специфичность и отрицательная прогностическая ценность (TNR) 96,5% и 98,1% соответственно для цоДНК позволяет говорить о возможности применения платформы для данного типа анализа даже в условиях повышенного уровня артефактов.

Детальное изучение методов обработки каждого типа данных позволяет предлагать способы увеличения числа биомаркерных событий для анализа в клинике. Так за счет объединения находок цоДНК и РНК тканевой биопсии с сырыми данными ДНК тканевой биопсии число пациентов с детектируемыми биомаркерными событиями возросло с 51 до 58. Широкое применение подобных методов требует дополнительной клинической валидации.

Выводы. Прямое пересечение клинически значимых находок показывает низкую перекрываемость между тканевой и жидкостной биопсиями (27%) и сниженную воспроизводимость между РНК и ДНК тканевой биопсии (пересечение 82%, корреляция аллельных частот 0,59), что согласуется с исследованиями в области. При этом объединение данных позволяет увеличить число пациентов с обнаруженными клинически значимыми событиями на 13%. Проанализированные методы комплексного геномного профилирования взаимодополняют друг друга, но в силу особенностей каждой из тестовых платформ не могут полностью заменить один другого. Настоящее исследование показывает клиническую необходимость интеграции данных секвенирования из разных источников.

Материалы и методы. Когорта 100 пациентов с разрешением к использованию данных в исследованиях: данные ДНК и РНК секвенирования тканевой биопсии (источник – формалиновые блоки), ДНК секвенирования жидкостной биопсии (источник – плазма крови), а также образцы нормальных тканей в качестве контроля врожденных мутаций. Секвенирование проводилось на платформе Illumina NovaSeq 6000. Распределение по диагнозам: саркома 37%, рак шейки матки 10%, немелкоклеточный рак легкого 9%, колоректальный рак 8%, меланома 6%, рак поджелудочной железы 6% и другие диагнозы солидных опухолей.

Список клинически значимых мутаций – Memorial Sloan Kettering Cancer Center [25]

Этапы биоинформатической обработки данных: выравнивание прочтений на референсный геном человека (hg38) с помощью Burrows-Wheeler Aligner (v0.7.17) [26] для данных ДНК секвенирования и с помощью STAR (v.2.4.2) [27] для данных РНК секвенирования, алгоритмы дедупликации и снижения шума, поиск мутаций в данных ДНК тканевой биопсии с помощью Strelka (v2.9.10) [28], в данных РНК секвенирования – PISCES (v5.10) [29], в данных ДНК жидкостной биопсии – LoFreq (v2.1.4) [30].

“BostonGene Technologies” ООО
anastasiya.yudina@bostongene.com

А.С. Юдина

Интеграция NGS данных комплексного геномного профилирования солидной и жидкостной биопсии для улучшения молекулярной диагностики онкологических заболеваний

Таргетная терапия онкологических заболеваний показала свою повышенную эффективность для ряда диагнозов относительно традиционных методов лечения. Для выбора и назначения таргетных терапий методы молекулярной диагностики должны соответствовать высоким стандартам чувствительности и специфичности получаемых результатов. В эру развития методов обработки и анализа больших данных перспективным подходом становится интеграция данных NGS от разных платформ и тканей. Наиболее распространены экстракция ДНК напрямую из опухолевой ткани (тканевая биопсия) и экстракция внеклеточной ДНК из плазмы (жидкостная биопсия). Молекулярные свойства, а также репертуар молекулярных событий в обеих тканях могут сильно различаться.

В настоящем исследовании было показано, что объединение соматических мутаций из тканевой и жидкостной биопсии с дополнительной про-

веркой находок в транскриптомном секвенировании позволило на 13% увеличить выявление клинически релевантных биомаркерных событий.

Ա.Ս. Յուդինա

**Պինդ և հեղուկ բիոպսիաների համապարփակ գենոմային
պրոֆիլավորումից ստացված NGS տվյալների ինտեգրում
քաղցկեղային հիվանդությունների մոլեկուլային ախտորոշման
բարելավման համար**

Քաղցկեղի թիրախային թերապիան ցույց է տվել ավելի բարձր արդյունավետություն մի շարք ուռուցքային հիվանդությունների դեպքում՝ բուժման ավանդական եղանակների համեմատ: Թիրախային բուժման ընտրության և նշանակման համար մոլեկուլային ախտորոշման մեթոդները պետք է համապատասխանեն ստացված արդյունքների զգայունության և սպեցիֆիկության բարձր չափանիշներին: Տվյալների մշակման և վերլուծության մեթոդների զարգացման դարաշրջանում խոստումնալից մոտեցում է հաջորդ սերնդի սեքվենավորման տվյալների ինտեգրումը տարբեր հարթակներից և հյուսվածքներից: Առավել տարածված են ԴՆԹ-ի անջատումը անմիջապես ուռուցքային հյուսվածքից (հյուսվածքային բիոպսիա) և արտաբջջային ԴՆԹ-ի անջատումը պլազմայից (հեղուկային բիոպսիա), որոնց դեպքում մոլեկուլային հատկությունները և գործընթացները կարող են էապես տարբեր լինել:

Այս աշխատանքում ցույց է տրվել, որ հյուսվածքային և հեղուկային բիոպսիայից ստմատիկ մուտացիաների տվյալների միավորումը՝ տրանսկրիպտոմային սեքվենավորմամբ արդյունքների լրացուցիչ ստուգման հետ մեկտեղ, 13%-ով ավելացրել է կլինիկորեն նշանակալի կենսամարկերների հայտնաբերումը:

A.S. Yudina

**Integration of NGS Data From Comprehensive Genomic Profiling of
Solid and Liquid Biopsies to Enhance Molecular Diagnostics for
Oncology Diseases**

Targeted therapy for cancer has demonstrated increased effectiveness for several diagnoses compared to traditional treatment methods. To select and prescribe targeted therapies, molecular diagnostic methods must meet high standards of sensitivity and specificity in the results obtained. In the era of big data processing and analysis, a promising approach is the integration of NGS data from different platforms and tissues. The most common methods involve DNA

extraction directly from tumor tissue (tissue biopsy) and the extraction of cell-free DNA from plasma (liquid biopsy). The molecular properties and the repertoire of molecular events in these two tissue types can vary significantly.

The present study showed that combining somatic mutations from tissue and liquid biopsies, with additional verification of findings in transcriptome sequencing, increased the clinically relevant biomarker events by 13%.

Литература

1. *Ismail R.K., Suijkerbuijk K.P.M., de Boer A. et al.* – Melanoma Research, 2022, V. 1, № 32(6), p. 460–8, <https://doi.org/10.1097/CMR.0000000000000832>.
2. *Aguilar A., Mas L., Enríquez D., et al.* – Cancer Control, 2022, V. 29, № 10732748211068637, pp. 1–8, <https://doi.org/10.1177/10732748211068637>.
3. *Alessandrini L., Perin T., Kadare S., et al.* – Current Cancer Drug Targets, 2018, V. 18, № 5, p. 410–20, <https://doi.org/10.2174/1568009618666171129145703>.
4. *American Society of Clinical Oncology (ASCO).* Understanding targeted therapy. Accessed at <https://www.cancer.net/navigating-cancer-care/how-cancer-treated/personalized-and-targeted-therapies/understanding-targeted-therapy> on December 19, 2019.
5. *Kris M.G., Johnson B.E., Berry L.D. et al.* – JAMA, 2014, V. 311, № 19, pp. 1998–2006, <https://doi.org/10.1001/jama.2014.3741>.
6. *Meric-Bernstam F., Brusco L., Shaw K. et al.* – Journal of Clinical Oncology, 2015, V. 33, № 25, pp. 2753–62, <https://doi.org/10.1200/JCO.2014.60.4165>.
7. *Palmeri M., Mehnert J., Silk A.W. et al.* – ESMO Open, 2022, V. 7, № 1, p. 100336, <https://doi.org/10.1016/j.esmoop.2021.100336>.
8. *Schrock A.B., Ouyang C., Sandhu J. et al.* – Annals of Oncology, 2019, V. 30, № 7, pp. 1096–103, <https://doi.org/10.1093/annonc/mdz134>.
9. *Keller L., Belloum Y., Wikman H. et al.* – British Journal of Cancer, 2021, V. 124, № 2, pp. 345–358, <https://doi.org/10.1038/s41416-020-01047-5>.
10. *Wen X., Pu H., Liu Q. et al.* – Cancers (Basel), 2022, V. 14, № 24, p. 6025, <https://doi.org/10.3390/cancers14246025>.
11. *Hahn A. W., Gill D. M., Maughan B. et al.* – Oncotarget, 2017, V. 8, № 20, pp. 33614–20, <https://doi.org/10.18632/oncotarget.16833>.
12. *Toor O.M., Ahmed Z., Bahaj W. et al.* – Molecular Cancer Therapeutics, 2018, V. 17, № 5, pp. 1123–32, <https://doi.org/10.1158/1535-7163.MCT-17-1015>.
13. *Chang Y.S., Fang H.Y., Hung Y.C. et al.* – Journal of Cancer Research and Clinical Oncology. 2018, V. 144, № 11, pp. 2167–75, <https://doi.org/10.1007/s00432-018-2747-9>.
14. *Park S., Olsen S., Ku B.M. et al.* – Cancer, 2021, V. 127, № 16, pp. 3019–3028, <https://doi.org/10.1002/cncr.33571>.
15. *Rodon Font N., No Garbarino Y., Díaz Castello O. et al.* – Revista Española de Patología, 2022, V. 55, № 3, pp. 156–62, <https://doi.org/10.1016/j.patol.2022.01.001>.

16. Guo Q., Lakatos E., Bakir I. A. *et al.* – Nature Communications, 2022, V. 13, № 1 p. 4487, <https://doi.org/10.1038/s41467-022-32041-5>.
17. Bhagwate A.V., Liu Y., Winham S.J. *et al.* – BMC Genomics, 2019, V. 20, № 1, p. 689, <https://doi.org/10.1186/s12864-019-6056-8>.
18. Gong B., Deveson I.W., Mercer T. *et al.* – Scientific Data, 2022, V. 9, № 1, p. 170, <https://doi.org/10.1038/s41597-022-01276-8>.
19. Koboldt D.C. – Genome Medicine, 2020, V. 12, № 1, p. 91, <https://doi.org/10.1186/s13073-020-00791-w>.
20. Castel S.E., Aguet F., Mohammadi P. *et al.* – Genome Biology, 2020, V. 21, № 1, p. 234 <https://doi.org/10.1186/s13059-020-02122-z>.
21. Gandara D.R., Paul S.M., Kowanetz M. *et al.* – Nature Medicine, 2018, V. 24, № 9, pp. 1441–8, <https://doi.org/10.1038/s41591-018-0134-3>.
22. Ma Y., Li Q., Du Y. *et al.* – Frontiers in Oncology, 2021, V. 11, p. 640761, <https://doi.org/10.3389/fonc.2021.640761>.
23. Kim E.S., Velcheti V., Mekhail T. *et al.* – Nature Medicine, 2022, V. 28, № 5, pp. 939–45, <https://doi.org/10.1038/s41591-022-01754-x>.
24. Atezolizumab – NCI. Accessed at <https://www.cancer.gov/about-cancer/treatment/drugs/atezolizumab> (2016).
25. Accessed at <https://www.cancerhotspots.org/#/home>.
26. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. – arXiv: Genomics.1303.3997v2 [q-bio.GN]. 2013.
27. Dobin A., Davis C.A., Schlesinger F. *et al.* – Bioinformatics, 2013, V. 29, № 1, pp. 15–21, <https://doi.org/10.1093/bioinformatics/bts635>.
28. Kim S., Scheffler K., Halpern A.L. *et al.* – Nature Methods, 2018, V. 15, № 8, pp. 591–4, <https://doi.org/10.1038/s41592-018-0051-x>.
29. Dunn T., Berry G., Emig-Agius D. *et al.* – Bioinformatics, 2019, V. 35, № 9, pp. 1579–81. <https://doi.org/10.1093/bioinformatics/bty849>.
30. Wilm A., Aw P.P., Bertrand D. *et al.* – Nucleic Acids Research, 2012, V. 40, № 22, pp. 11189–201, <https://doi.org/10.1093/nar/gks918>.