

УДК 57.05, 57.07

Эпигенетика древней ДНК

С. В. Женило, А. С. Соколов, Е. Б. Прохорчук*

Институт биоинженерии, Федеральный исследовательский центр «Фундаментальные основы биотехнологии» РАН, 117312, Москва, просп. 60-летия Октября, 7, корп. 1

*E-mail: prokhortchouk@gmail.com

Поступила в редакцию 29.01.2016

Принята к печати 25.04.2016

РЕФЕРАТ Изначально работа с древней ДНК была основана на анализе ее нуклеотидной последовательности. Этот подход позволяет изучать эволюционные изменения, происходящие в различных популяциях, определять влияние окружающей среды на генетический отбор. Однако усовершенствование методических подходов к проведению полногеномного анализа открыло новые возможности изучения эпигенетических механизмов, вовлеченных в регуляцию экспрессии генов. Один из наиболее ярких примеров влияния эпигенетических модификаций на эволюцию человека – изменение статуса метилирования регуляторных последовательностей целого ряда генов кластера *HOXD* и гена *MEIS1*. Эпигенетические изменения в этих генах играют ключевую роль в эволюции конечностей современного человека. Последние работы показали возможность определения транскрипционной активности генов в образцах древней ДНК путем комбинирования информации о метилировании ДНК и о наличии гиперчувствительных к ДНКазе I участков в последовательностях, расположенных в стартах транскрипции генов. В дальнейшем при нахождении хорошо сохранившихся тканей появляется перспектива обнаружения эволюционных изменений, связанных с эпигенетическими различиями в высшей нервной деятельности современного и древнего человека.

КЛЮЧЕВЫЕ СЛОВА древняя ДНК, метилирование ДНК, эпигенетика.

СПИСОК СОКРАЩЕНИЙ МрG – метилированные CpG; PMD – частично метилированные домены (partially methylated domains); DMV – «долины» метилированной ДНК (DNA methylation valley); MBD – метил-ДНК-связывающий домен (methyl DNA binding domain).

ВВЕДЕНИЕ

Изучение ДНК, выделенной из археологических и палеонтологических образцов, позволяет получить информацию об эволюционном прошлом. Первое время работа с древней ДНК заключалась в анализе нуклеотидных последовательностей. На этом этапе исследователи сталкивались с множеством трудностей, связанных с качеством древней ДНК, ее контаминацией чужеродной ДНК и пр. Эти сложности удалось преодолеть, в частности, совершенствованием методов полногеномного секвенирования. Однако развитие современных технологий секвенирования позволяет также анализировать информацию, содержащуюся в надгенетическом коде. С одной стороны, физические характеристики, предрасположенность к заболеваниям и даже некоторые психологические особенности индивида обусловлены генетическими факторами. С другой стороны, нельзя забывать про воздействие окружающей среды. Экспрессия генов определяется не только нуклеотидной последовательностью, но и целым рядом адаптационно регулируемых процессов, которые приводят к изменению уровня метилирования ДНК, гистонового кода, спектра

микроРНК. Эти эпигенетические механизмы вовлечены в формирование структур хроматина, необходимых для регуляции экспрессии генов. Благодаря комбинации высокотехнологического секвенирования и различных методических подходов получены карты полногеномного метилирования ДНК в различного типа клетках человека и мыши: соматических, стволовых, половых, раковых и т.д. [1].

Основной характеристикой любой древней ДНК является ее деградация и дезаминирование остатков цитозина. До недавнего времени считалось невозможным получить информацию о транскрипционной активности генов на основе ДНК, выделенной через длительное время после смерти индивида. Тем не менее в 2010 году группой С. Паабо были предприняты первые попытки получения карт метилирования древней ДНК, которые показали потенциальную возможность определения картины метилирования CpG *in vivo* в ДНК неандертальца [2]. Немного позже с помощью бисульфитного аллель-специфичного секвенирования отдельных локусов ДНК бизона из образцов, относящихся к позднему плейстоцену, Б. Лламас и соавт. [3] показали, что в древней ДНК сохраняется метилирование.

МЕТОДЫ ИССЛЕДОВАНИЯ МЕТИЛИРОВАНИЯ ДРЕВНЕЙ ДНК

Дезаминирование метилированных остатков цитозина и их превращение в тимин, происходящее после смерти, затрудняет количественный анализ метилированных цитозинов. Только в 2014 году был разработан метод, позволяющий проводить полногеномный анализ метилирования древней ДНК [4–6]. При бисульфитном секвенировании неметилированные остатки цитозина химически конвертируются в остатки урацила, которые затем считываются полимеразой, например Taq-полимеразой, как тимины при ПЦР. В клетках позвоночных картирование таких С-Т-мутаций осуществляется с точностью до одного нуклеотида. Аналогичное химическое превращение происходит после смерти в естественных условиях в основном за счет гидролитического дезаминирования цитозинов, расположенных в одноцепочечных участках [7]. Применение Taq-полимеразы, например Taq platinum high fidelity (Hifi), нечувствительной к присутствию урацила, приводит к увеличению количества С-Т-замен относительно исходной цепи. Такие замены не могут наблюдаться при использовании такой Pfu-полимеразы, как Phusion. Pfu-полимераза не способна продолжать синтез цепи в присутствии урацила. При этом 5-метилцитозин дезаминируется в тимин, в отличие от немодифицированного цитозина, и может быть успешно амплифицирован Pfu-полимеразой (рис. 1). Таким образом, увеличение количества С-Т-замен при анализе древней ДНК позволяет отличить метилированные цитозины от неметилированных [5]. Дезаминирование – это стохастический процесс, значит, всегда есть метилированные CpG-динуклеотиды (Mpg), не дезаминированные с течением времени, которые будут амплифицироваться как обычные CpG (рис. 1). Поэтому данный метод позволяет определить метилирование не всех цитозинов, а только дезаминированных. Для определения метилированных цитозинов с однонуклеотидным разрешением необходимо увеличивать уровень покрытия. Так, в работе группы Р. Смита [8] показана возможность определять статус метилирования одиночного CpG-динуклеотида, расположенного в мобильном генетическом элементе LINE-1, в образцах древней ДНК, полученной из скелетов пяти североамериканцев, возраст которых составлял от 200 до 4000 лет до нашей эры. Если не требуется определять Mpg с точностью до одного нуклеотида, то достаточно анализировать количество CpG-ТрG-замен на интересующем участке ДНК. Такой подход применен к анализу генома неандертальского, денисовского, палеоэскимосского человека [4]. Удалось

показать, что в костном мозге и в волосяном фолликуле современных и древних людей метилирование ДНК практически идентично.

АНАЛИЗ ЭПИГЕНОМА ДРЕВНЕЙ ДНК

При исследовании дифференциально метилированных участков Д. Джохман и соавт. обнаружили, что промоторы и последовательности генов *HOXD9* и *HOXD10*, одних из ключевых регуляторов развития конечностей, метилированы в костном мозге древних образцов (человек неандертальский и денисовский) и неметилированы в ДНК современного человека [4]. На мышинных модельных системах показано, что изменения в экспрессии генов кластера *HOXD*, особенно генов *HOXD9* и *HOXD10*, приводят к морфологическим изменениям [9], которые напоминают различия в строении конечностей неандертальца и современного человека. Этот факт позволяет предположить, что эпигенетические изменения в кластере генов *HOXD* играют ключевую роль в эволюции конечностей современного человека. Более того, внутри гена *MEIS1*, кодирующего белок, который регулирует активность генов кластера *HOXD*, найден дифференциально метилированный участок [4]. Информация о метилировании древней ДНК больших геномных районов в виде количества С-Т-замен может использоваться для поиска протяженных участков с измененным уровнем метилирования в костном мозге и волосяных фолликулах наших предков. Такой анализ позволит обнаружить не только гиперметилированные CpG-островки, но и:

1) длинные (от нескольких сотен тысяч до нескольких миллионов пар нуклеотидов) частично метилированные домены (PMD), которые не содержат гены и колокализуются с ламинсвязанными структурами [10, 11];

2) «долины» метилированной ДНК (DMV) (несколько тысяч пар нуклеотидов), гипометилированные в большинстве тканей, содержащие большое количество генов развития и участков связывания факторов транскрипции [12, 13], но гиперметилированы в клетках рака кишечника [12];

3) неометилированные «каньоны» (десятки тысяч пар нуклеотидов), недавно открытые в гемопоэтических стволовых клетках [14];

4) эпигенетические механизмы воспаления кишечника, которые характеризуются гиперметилированием DMV, низкой плотностью CpG, активными марками хроматина [15].

Все перечисленные исследования основаны на анализе уровня метилирования геномных областей, длина которых варьирует от нескольких тысяч до нескольких миллионов пар нуклеотидов. Эпигенетический анализ древней ДНК, базирующийся-

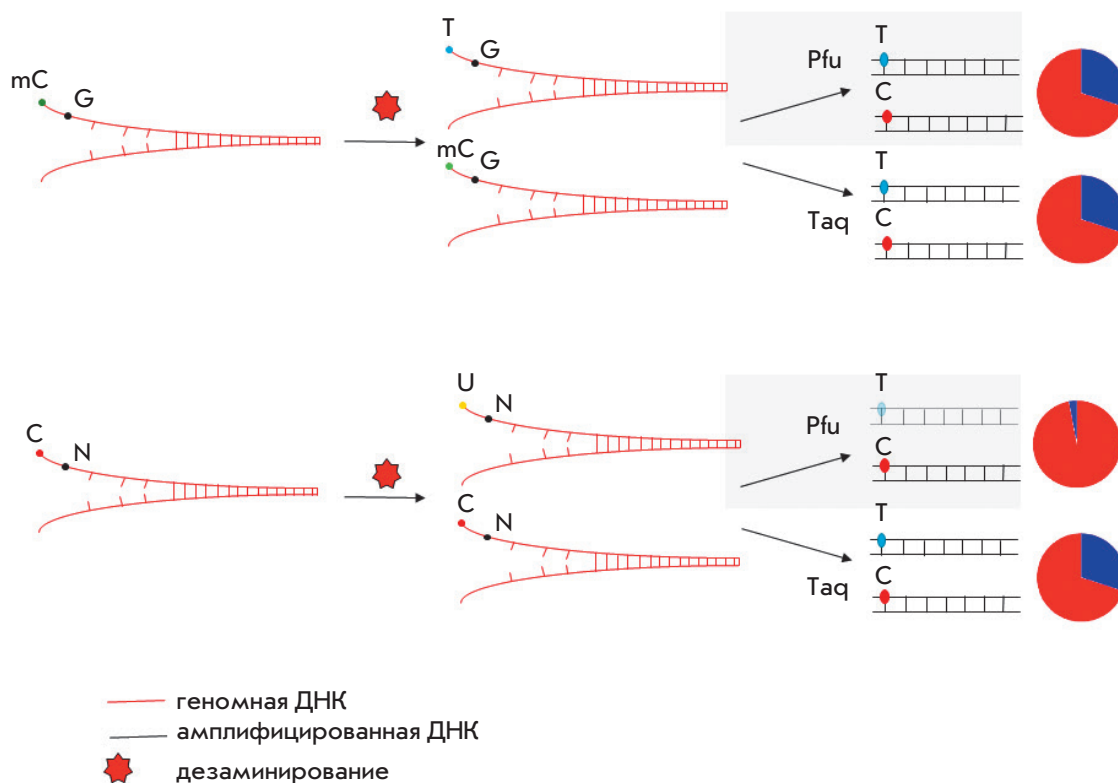


Рис. 1. Схема определения метилирования цитозина в древней ДНК за счет увеличения количества С-Т-замен относительно исходной цепи при дезаминировании. Таq-полимераза не чувствительна к урацилу, а Pfu-полимераза не может продолжать синтез цепи в присутствии урацила. Дезаминирование метилированных цитозинов (mC) не отразится на соотношении С-Т-замен, выявляемых с помощью Таq- и Pfu-полимераз. В случае неметилированных цитозинов дезаминирование создает «непроходимый» для Pfu-полимеразы урацил. Таким образом, все чтения, которые начинаются на Т (в контексте геномного CG), в NGS-библиотеках, приготовленных с использованием Pfu-полимеразы, однозначно говорят о метилированном состоянии соответствующего цитозина в древней ДНК

ся на поиске регионов, обогащенных заменами С-Т, открывает перед нами новые возможности определения адаптационных сигналов и/или маркеров заболеваний. Однако для этого необходимы хорошо сохранившиеся ткани (мозг, кишечник, мышцы, кровь), что характерно для останков, найденных в вечной мерзлоте, например, останков мамонтов, живших в эпоху плейстоцена (см. далее). В 2014 году была опубликована уникальная работа по составлению эпигенетической карты, полученной на ДНК одного волосяного фолликула палеоэскимосского человека, позволившая определить возраст индивида в момент смерти [5]. Это сделано на основе судебного исследования, показавшего возможность вычисления возраста по уровню метилирования определенных CpG-динуклеотидов [16]. Если предположить, что 6000 лет назад внешняя среда воздействовала на метилирование так же, как в наши дни, то, основываясь на со-

временных базах данных, можно определить возраст древнего человека. Группа Л. Орландо показала, что саккакский человек (палеоэскимосская эпоха в Гренландии) был сравнительно немолод, к моменту смерти ему, вероятно, было около 35–40 лет [5].

ПРОБЛЕМА КОНТАМИНАЦИИ ДРЕВНЕЙ ДНК

Значительную трудность при работе с древней ДНК представляет контаминация материала бактериальной ДНК. Недавно показали, что эпигенетические особенности позвоночных (метилирование CpG) можно применить для разделения бактериальной и древней геномной ДНК человека [17]. Метилированные CpG-динуклеотиды найдены только в соматических клетках позвоночных. В бактериальном геноме также есть метилированные цитозины и аденины, но не в CpG. В состав белков семейства MBD входит метил-ДНК-связывающий домен (MBD),

который взаимодействует с метилированной ДНК, содержащей одиночные метилированные CpG [18]. Аффинная хроматография на основе домена MBD является рутинным методом построения карт метилирования геномов различных организмов. Этот метод позволяет не только охарактеризовать древние метиломы, но и разделить ДНК древних позвоночных и микроорганизмов. На примере палеоэскимосского саккакского человека, шерстистых мамонтов (Юка и Хрома), полярных медведей и двух видов, принадлежащих к семейству лошадиных, показано, что метилирование сохраняется в различных тканях, при внешних воздействиях и при разном времени появления останков (от 4000 до 45000 лет до нашего времени). Обогащение с помощью аффинной хроматографии с MBD-доменом позволяет проводить анализ микробиоты древних образцов, а также потенциально патогенных геномов [17].

ТРАНСКРИПЦИОННАЯ АКТИВНОСТЬ ДРЕВНИХ ГЕНОВ НА ОСНОВЕ НУКЛЕОСОМНЫХ КАРТ

Метилирование ДНК может служить маркером подавления транскрипции генов, однако этой информации недостаточно для того, чтобы утверждать, экспрессировался ген или нет. Для предсказания транскрипционной активности гена необходима дополнительная информация, например, о модификациях гистонов, структуре хроматина, связывании факторов транскрипции с регуляторными участками. Предприняты первые попытки работы с белками из древних организмов [19]. Секвенирование древней ДНК выявило неожиданный источник эпигенетической информации. Исследователи из группы Л. Орландо обнаружили периодичность в плотности прочтения нуклеотидов [5]. Они предположили, что наблюдаемая периодичность не является ошибкой выравнивания последовательностей или артефактом секвенирования. В основе периодичности плотности прочтения нуклеотидов лежит защита ДНК путем связывания ее с нуклеосомами, при этом расщепление ДНКазой, которая попала в ядро во время клеточной смерти, либо посмертная деградация ДНК происходят в линкерных участках. В этом случае наблюдаемая глубина секвенирования отражает положение нуклеосом на ДНК. Спектральные методы анализа ДНК применяются для поиска скрытых периодичностей. Так, в случае сравнительно коротких последовательностей использование преобразования Фурье позволяет получать статистические критерии, обладающие свойством самоусредняемости [20]. Применение преобразования Фурье к функции, которая связывает плотность покрытия «чтениями» с координатой в геноме, имеет сильный пик на 180–190 нуклеотидов, указывая на то, что пе-

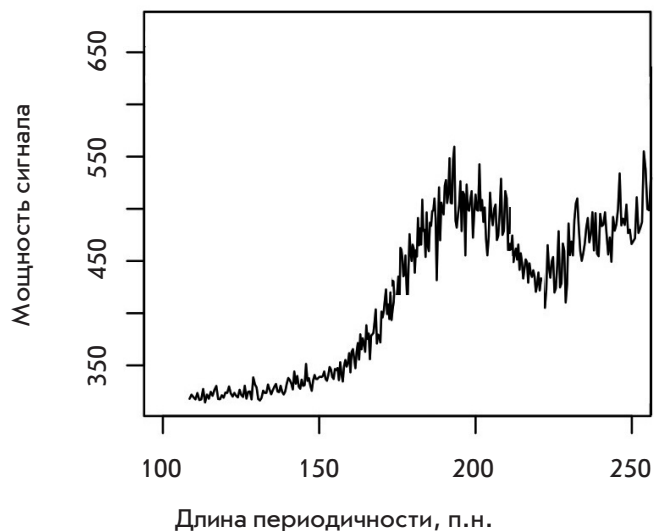


Рис. 2. График спектральной плотности (оценка спектральной мощности). Анализ регионов вокруг стартов транскрипции (+/- 1000 п.н.) генов, проведена коррекция относительно фонового распределения чтений при анализе современной ДНК слона. Используются данные секвенирования генома детеныша мамонта, найденного в 2009 году в нижнем течении реки Хрома (Прохорчук Е.Б. и соавт, неопубликованные данные). Возраст находки более 50000 лет

риодичность покрытия при секвенировании совпадает с периодичностью организованного в нуклеосомы хроматина (рис. 2). При анализе распределения плотности расположения 5'-конца чтений выяснилось, что на расстояниях с характерной длиной 100 нуклеотидов проявляется также периодичность в 10 нуклеотидов, которую связывают с одним оборотом спирали ДНК: нуклеотиды, обращенные к нуклеосомам, не будут стартом чтений, поскольку они менее доступны для нуклеаз. В древней ДНК с высокой точностью удалось картировать положение нуклеосом на расстоянии 4 т.п.н. от сайтов связывания СТСФ, расположение которых отрицательно коррелировало с метилированием ДНК. Активно транскрибируемые участки можно находить с помощью анализа гиперчувствительных к ДНКазе I участков, расположенных в стартах транскрипции [21]. Предполагается, что участки с открытой структурой хроматина будут более чувствительны к расщеплению ДНКазой I при апоптозе или после смерти организма. Поэтому при полногеномном секвенировании плотность чтений в районах старта транскрипции активных генов будет снижена по сравнению с молчащими генами. Если одновременно учитывать плотность чтений в области определенного старта транскрипции

и информацию о гиперчувствительных к ДНКазе I участках из базы данных ENCODE, то появляется потенциальная возможность определения транскрипционной активности соответствующих генов.

ПЕРСПЕКТИВЫ

Комбинирование информации о метилировании ДНК и сайтах гиперчувствительности к ДНКазе I в стартах транскрипции позволит в ближайшее время восстановить количественную картину экспрессии

генов в древних образцах. Если удастся найти хорошо сохранившийся мозг древнего человека, то это будет прорывом в области изучения древней ДНК и эволюции человека. Подобную возможность подтверждает недавняя находка шерстистого мамонта, структуры мозга которого хорошо сохранились [22]. Предполагается, что основные отличия в высшей нервной деятельности древних и современных людей будут найдены на эпигенетическом уровне [23, 24]. ●

СПИСОК ЛИТЕРАТУРЫ

- Gerstein M.B., Kundaje A., Hariharan M., Landt S.G., Yan K.K., Cheng C., Mu X.J., Khurana E., Rozowsky J., Alexander R., et al. // *Nature*. 2012. V. 489. P. 91–100.
- Briggs A.W., Stenzel U., Meyer M., Krause J., Kircher M., Paabo S. // *Nucl. Acids Res.* 2010. V. 38. P. e87.
- Llamas B., Holland M.L., Chen K., Cropley J.E., Cooper A., Suter C.M. // *PLoS One*. 2012. V. 7. P. e30226.
- Gokhman D., Lavi E., Prufer K., Fraga M.F., Riancho J.A., Kelso J., Paabo S., Meshorer E., Carmel L. // *Science*. 2014. V. 344. P. 523–527.
- Pedersen J.S., Valen E., Velazquez A.M., Parker B.J., Rasmussen M., Lindgreen S., Lilje B., Tobin D.J., Kelly T.K., Vang S., et al. // *Genome Res.* 2014. V. 24. P. 454–466.
- Smith O., Clapham A.J., Rose P., Liu Y., Wang J., Allaby R.G. // *Sci Rep.* 2014. V. 4. P. 5559.
- Hofreiter M., Jaenicke V., Serre D., von Haeseler A., Paabo S. // *Nucleic Acids Res.* 2001. V. 29. P. 4793–4799.
- Smith R.W., Monroe C., Bolnick D.A. // *PLoS One*. 2015. V. 10. P. e0125344.
- Zakany J., Duboule D. // *Curr. Opin. Genet. Dev.* 2007. V. 17. P. 359–366.
- Hansen K.D., Timp W., Bravo H.C., Sabunciyan S., Langmead B., McDonald O.G., Wen B., Wu H., Liu Y., Diep D., et al. // *Nat. Genet.* 2011. V. 43. P. 768–775.
- Berman B.P., Weisenberger D.J., Aman J.F., Hinoue T., Ramjan Z., Liu Y., Noushmehr H., Lange C.P., van Dijk, C.M., Tollenaar R.A., et al. // *Nat. Genet.* 2012. V. 44. P. 40–46.
- Xie W., Schultz M.D., Lister R., Hou Z., Rajagopal N., Ray P., Whitaker J.W., Tian S., Hawkins R.D., Leung D., et al. // *Cell*. 2013. V. 153. P. 1134–1148.
- Nakamura R., Tsukahara T., Qu W., Ichikawa K., Otsuka T., Ogoshi K., Saito T.L., Matsushima K., Sugano S., Hashimoto S., et al. // *Development*. 2014. V. 141. P. 2568–2580.
- Jeong M., Sun D., Luo M., Huang Y., Challen G.A., Rodriguez B., Zhang X., Chavez L., Wang H., Hannah R., et al. // *Nat. Genet.* 2014. V. 46. P. 17–23.
- Abu-Remaileh M., Bender S., Raddatz G., Ansari I., Cohen D., Gutekunst J., Musch T., Linhart H., Breiling A., Pikarsky E., et al. // *Cancer Res.* 2015. V. 75 (10). P. 2120–2130.
- Kader F., Ghai M. // *Forensic Sci. Int.* 2015. V. 249. P. 255–265.
- Seguin-Orlando A., Gamba C., Sarkissian C.D., Ermini L., Louvel G., Boulygina E., Sokolov A., Nedoluzhko A., Lorenzen E.D., Lopez P., et al. // *Sci. Repts.* 2015. Published online number: 11826.
- Hendrich B., Bird A. // *Mol. Cell Biol.* 1998. V. 18. P. 6538–6547.
- Welker F., Collins M.J., Thomas J.A., Wadsley M., Brace S., Cappellini E., Turvey S.T., Reguero M., Gelfo J.N., Kramarz A., et al. // *Nature*. 2015. V. 522. P. 81–84.
- Лобзин В.В., Чечеткин В.П. // *Успехи физ. наук*. 2000. Т. 43. С. 55–78.
- Crawford G.E., Holt I.E., Whittle J., Webb B.D., Tai D., Davis S., Margulies E.H., Chen Y., Bernat J.A., Ginsburg D., et al. // *Genome Res.* 2006. V. 16. P. 123–131.
- Kharlamova A., Kurtova A., Chernikov V., Protopopov A., Boeskorov G., Plotnikov V., Ushakov V., Maschenko E. // *Quaternary International*. 2016. P. 86–93.
- Zeng J., Konopka G., Hunt B.G., Preuss T.M., Geschwind D., Yi S.V. // *Am. J. Hum. Genet.* 2012. V. 91. P. 455–465.
- Chopra P., Papale L.A., White A.T., Hatch A., Brown R.M., Garthwaite M.A., Roseboom P.H., Golos T.G., Warren S.T., Alisch R.S. // *BMC Genomics*. 2014. V. 15. P. 131.