

УДК 616.98

ТОЧНОСТЬ ОПРЕДЕЛЕНИЯ СУБТИПА ВИЧ-1 НА ОСНОВАНИИ АНАЛИЗА НУКЛЕОТИДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ V3 ПЕТЛИ ГЕНА *gp120*

М.Ю.Дмитрюкова, Д.Е.Киреев, А.Э.Лопатухин, И.А.Лаповок, Г.А.Шипулин
Центральный НИИ Эпидемиологии, Москва, Россия

ACCURACY OF HIV-1 SUBTYPE ASSIGNMENT OF V3 LOOP GP120 GENE

M.Y.Dmitryukova, D.E.Kireev, A.E.Lopatukhin, I.A.Lapovok, G.A.Shipulin
Central Research Institute for Epidemiology, Moscow, Russia

© Коллектив авторов, 2015 г.

Введение в рутинную лабораторную практику в России секвенирования гена *gp120* белка оболочки ВИЧ-1, необходимого перед назначением препарата маравирик, относящегося к классу CCR5-антагонистов, позволит накопить достаточное количество нуклеотидных последовательностей региона *env*. Эта информация может оказаться полезной не только для назначения эффективной схемы терапии, но и для эпидемиологического надзора. «Золотым» стандартом при определении субтипа вируса является филогенетический метод, однако в практике в силу ряда сложностей он используется довольно редко. Значительно чаще применяется автоматическое субтипирование с использованием специальных программ, основанных на различных статистических или математических подходах. В настоящее время достоверность определения субтипа ВИЧ-1 при анализе нуклеотидных последовательностей V3 петли с помощью автоматических алгоритмов вариантов ВИЧ, циркулирующих в России, изучена недостаточно. В этой работе мы определили, насколько точно анализ фрагмента гена *gp120*, расположенного в регионе *env*, позволяет определить субтип ВИЧ-1. Совпадение результатов автоматического субтипирования по регионам *pol* и *env* составило 84,7%. Дискордантные результаты были связаны в первую очередь, с недостаточной чувствительностью в отношении рекомбинантных вариантов ВИЧ-1 при субтипировании по региону *env*, а также некорректной работой алгоритмов автоматического субтипирования.

Ключевые слова: ВИЧ-1, V3 петля, *gp120*, регион *env*, субтипирование, *geno2pheno*, REGA.

The use of HIV-1 genotyping for tropism prediction before Maraviroc (MVC, Pfizer) prescription in Russia will produce a large number of *env* gene sequences. These data may be useful for both therapy and epidemiological surveillance. But HIV-1 subtyping tools performance on Russian strains is still undetermined. In this work we have investigated the accuracy of genotypic subtype prediction by Geno2pheno algorithm on *env* gene region. Concordant results between *pol* and *env* regions were for 84,7% samples. Agreement between phylogenetic analysis and automatic prediction by Geno2Pheno of *env* gene region was found in 90,6% that means quite high accuracy of the algorithm. Discordant results were associated with low sensitivity for recombinants and inaccurate algorithm performance.

Key words: HIV-1, *env* gene, subtyping, Geno2pheno, REGA.

Введение. Эпидемия ВИЧ-инфекции на сегодняшний день охватывает почти 35 млн. людей во всем мире [1]. В Российской Федерации количество людей, живущих с диагнозом «ВИЧ-инфекция», приближается к 750 тыс. человек [2].

Вирус иммунодефицита человека представляет собой гетерогенную группу, в которой выделяют два типа вируса, ВИЧ-1 и ВИЧ-2, при этом эпидемически значимым является ВИЧ типа 1. ВИЧ-1 также неоднороден и включает 4 группы (M, N, O, P). Наибольшее распространение получила группа M, отвечающая более чем за 90% случаев инфекции. В ней на данный момент выделяют 9 субтипов ВИЧ

(A-D, F-H, J-K) и, по крайней мере, 61 вариант циркулирующих рекомбинантных форм (circulating recombinant form, CRF) [3]. Кроме того, обнаруживается множество уникальных рекомбинантных форм (URF, unique recombinant form), пока не отвечающих требованиям включения в CRF (известная последовательность, наличие трех не связанных между собой случаев инфицирования) [4]. Распределение различных субтипов и вариантов в мире неоднородно. В Российской Федерации наибольшее распространение (до 85%) получил генетический вариант IDU-A, субтипа A, в меньшей степени субтип B, рекомбинант CRF03_AB и некоторые другие [5].

Наблюдение за генетическим разнообразием вариантов вируса в различных регионах мира ведется с самого начала установления надзора за эпидемией. Это необходимо в первую очередь для установления эпидемиологических связей. Кроме того, от субтипа вируса зависят точность работы скрининговых тестов и тестов для определения вирусной нагрузки, скорость прогрессирования ВИЧ-инфекции и вероятность возникновения устойчивости к антиретровирусным (АРВ) препаратам, в том числе смена тропизма вируса [6].

Наиболее эффективным методом определения субтипа вируса («золотым стандартом») является филогенетический анализ. Однако данный метод в силу относительно большой трудоемкости и сложности используется довольно редко. После начала широкомасштабного применения АРВ препаратов и затем тестов для определения лекарственной устойчивости на основе секвенирования, определение субтипа стало возможным в рутинной практике. Широкое распространение получили автоматические алгоритмы субтипирования, обычно сопряженные с алгоритмами определения лекарственной устойчивости. К их преимуществам относятся высокая скорость работы и простота использования. Поскольку в большинстве случаев определение устойчивости вируса к АРВ препаратам связано с секвенированием областей, кодирующих протеазу и обратную транскриптазу, расположенных в регионе *pol*, наиболее распространенные алгоритмы позволяют определять субтип на основании анализа именно этого фрагмента вирусного генома. Примерами могут послужить такие алгоритмы, как Stanford (<http://hivdb.stanford.edu/>), Geno2pheno (<http://www.geno2pheno.org/>), Comet (<http://comet.retrovirology.lu/>) и другие.

Кроме того, существует специально созданный для проведения субтипирования ВИЧ алгоритм REGA, курируемый университетом KU Leuven (Лейвен, Бельгия), (<http://dbpartners.stanford.edu:8080/RegaSubtyping/stanford-hiv/typing-tool>), использующий для этого филогенетический анализ [7]. Основное преимущество алгоритма REGA перед другими приложениями для субтипирования заключается в том, что он позволяет определять субтип по любому фрагменту вирусного генома. Остальные программы в большей степени основаны на сравнении имеющейся последовательности с последовательностями, собранными в их базе, в результате чего точность анализа зависит от того, насколько полно представлены те или иные субтипы или рекомбинантные формы в этой базе [8].

Субтипирование по нуклеотидным последовательностям региона *env* распространено меньше, так как

обширные исследования этой области в клинической практике начались сравнительно недавно — после разрешения к применению препарата маравирок (MVC, Pfizer), первого препарата класса ССR5-антагонистов. Использование этого препарата целесообразно только для пациентов, имеющих ССR5-тропный вариант ВИЧ, поэтому определение тропизма вируса необходимо для принятия решения о включении маравирока в состав АРВ терапии. Определение тропизма ВИЧ-1 проводят либо с использованием фенотипического метода (однако процедура довольно длительна и дорогостояща), либо с использованием генотипического метода, на основании анализа последовательности V3 петли гена *gp120* белка оболочки, расположенного в регионе *env*. Программа `geno2pheno[coreceptor]` (<http://coreceptor.bioinf.mpi-inf.mpg.de/index.php>), разработанная для генотипического определения тропизма ВИЧ, позволяет также определять субтип вируса по указанному фрагменту.

Целью данной работы явилась оценка точности определения субтипа ВИЧ-1 на основании анализа нуклеотидных последовательностей V3 петли гена *gp120*.

Материалы и методы исследования. В работе были использованы 85 образцов плазмы крови от ВИЧ-инфицированных лиц из различных регионов РФ (Московская, Ленинградская, Свердловская, Тверская, Нижегородская, Омская области; Татарстан, Хабаровский край и Красноярский край). Последовательности протяженностью приблизительно 1200 п.н. региона *pol* и 105 п.н. региона *env* (область V3 петли) были получены с использованием набора реагентов «АмплиСенс HIV-Resist-Seq» (ФБУН ЦНИИ Эпидемиологии Роспотребнадзора, Россия) в соответствии с инструкцией производителя в ходе рутинных исследований тропизма и лекарственной устойчивости вируса.

Филогенетический анализ последовательностей региона *pol* и V3 петли региона проводили методом ближайших соседей с использованием пакета программ MEGA6.06 [9]. Для создания древа были использованы 170 референсных последовательностей региона *env* и 160 последовательностей региона *pol*, депонированных в базе Национальной лаборатории Лос-Аламос, США (<http://hiv.lanl.gov>). Автоматическое субтипирование по области региона *pol* проводили с помощью алгоритмов REGA 3.0 [7] и Comet 1.0, по области региона *env* — с помощью алгоритмов `geno2pheno[coreceptor]` 2.5 [10] и Comet 1.0 [14].

Результаты исследования и их обсуждение. Результаты субтипирования образцов различными способами по разным регионам представлены в таблице 1.

По результатам автоматического субтипирования последовательностей региона *pol* в REGA 3.0 в исследованной группе было выделено 2 субтипа: А — 88,2%, В — 2,4%, а также рекомбинантные формы CRF03_AB — 5,9% и CRF02_AG — 3,5%. Проведение субтипирования ВИЧ с помощью алгоритма REGA по участку V3 петли региона *env* оказалось невозможным из-за его недостаточной протяженности.

программой к субтипам С, D и рекомбинанту 01_AE, соответственно. В результате автоматического субтипирования субтип А и рекомбинант 02_AG были объединены в одну группу (А or AG), таким образом, точное количество рекомбинанта 02_AG определить при субтипировании региона *env* в программе *geno2pheno* невозможно. Совпадение результатов субтипирования с помощью алгоритма *geno2pheno* по региону *env* с результатами филогене-

Таблица 1

Результаты субтипирования 85 образцов по регионам *pol* и *env*

Количество образцов	Регион <i>pol</i>		Регион <i>env</i>		
	REGA/ Comet	Филогенетический анализ	<i>geno2pheno</i>	Филогенетический анализ	Comet
67	A	A	A or AG	A	A1
1	A	A	A or AG	A	35_AD
2	A	A	B	B	B
2	A	A	B	03_AB	B
1	B	B	B	B	C (CRF29_BF)
1	B	B	A or AG	A	A1
2	03_AB	03_AB	B	03_AB	B (03_AB)
1	03_AB	03_AB	A	03_AB	D (06)
2	03_AB	03_AB	A or AG	A	A1
1	02_AG	02_AG	A or AG	02_AG	A1
1	02_AG	02_AG	A or AG	02_AG	35_AD
1	02_AG	63_02A1	A or AG	02_AG	A1
1	A	A	D	A	A1
1	A	A	C	A	A1
1	A	A	01_AE	A	A1

В результате филогенетического анализа региона *pol* образцы кластеризовались в пяти группах: субтипы А (88,2%) и В (2,4%), CRF03_AB (5,9%), 02_AG (2,4%). Кроме того, один образец кластеризовался с образцами варианта CRF63_02A1 (1,2%). Этот образец оказался единственным, результаты типирования которого отличались при различных способах анализа. Таким образом, результаты автоматического субтипирования по региону *pol* практически полностью (98,8%) совпали с результатами филогенетического анализа.

Филогенетический анализ нуклеотидных последовательностей по региону *env* при схожей кластеризации показал несколько иное распределение образцов: субтипы А (84,7%) и В (3,5%), CRF03_AB (5,9%), 02_AG (3,5%). Степень совпадения между результатами субтипирования с помощью филогенетического анализа по разным регионам составила 90,6%.

Работа программы *Geno2pheno* была оценена путем анализа последовательностей V3 петли. К субтипу А были отнесены 88,2% образцов, к субтипу В — 8,2%; еще три образца были отнесены

нетического анализа, сделанного по региону *env*, составило 90,6%, а по региону *pol* — 83,5%.

Так же была оценена работа программы Comet на указанной выборке образцов. Совпадение результатов автоматического субтипирования по фрагменту *pol* полностью совпало с результатами работы программы REGA. Совпадение результатов субтипирования с помощью алгоритма Comet по региону *env* с результатами филогенетического анализа, сделанного по региону *env*, составило 90,6%, а по региону *pol* — 84,7%. Таким образом, относительно референсного метода (филогенетического анализа) точность субтипирования на анализируемой выборке с помощью алгоритмов Comet и *geno2pheno* оказалась практически одинаковой. При этом между собой эти алгоритмы совпали только на 91,8%.

Расхождение результатов при определении субтипа по разным регионам вирусного генома, как с использованием автоматических программ, так и при проведении филогенетического анализа, связано, в первую очередь, с наличием рекомбинации, когда исследуемые фрагменты генома относятся

к разным субтипам вируса. В нашем исследовании дискордантные результаты такого рода связаны с рекомбинантами 03_AB и 02_AG.

Циркулирующая рекомбинантная форма 03_AB является наиболее распространенным рекомбинантом в Российской Федерации. Регион gag и часть региона pol этого рекомбинанта представлен фрагментом субтипа А, участок гена *gp120* — субтипом В [11]. В нашем случае образцы, в филогенетическом анализе по региону pol кластеризовавшиеся с последовательностями субтипа АВ, при автоматическом субтипировании имели результат А или АВ по региону pol, либо субтип В по региону env.

Схожая ситуация обстоит с циркулирующей рекомбинантной формой CRF02_AG, регион env

Подобного рода ошибки связаны с несовершенством алгоритмов субтипирования и встречаются достаточно часто [12, 13]. Связано это с особенностями эпидемии ВИЧ-инфекции в странах Западной Европы, где были впервые разработаны и валидированы программы определения устойчивости к лекарственным препаратам и субтипированию ВИЧ. Большинство случаев инфекции в этих странах вызвано субтипом В, для которого и показана наибольшая точность и воспроизводимость результатов работы алгоритмов. На так называемых не-В субтипах из-за недостаточного количества известных и проанализированных последовательностей расхождение может составлять до 25%, в зависимости от субтипа и используемого алгоритма [14].

Таблица 2

Сходимость различных методов субтипирования по разным регионам

	Регион pol, филогенетический анализ	Регион pol, REGA/Comet	Регион env, филогенетический анализ	Регион env, geno2pheno	Регион env, Comet
Регион pol, филогенетический анализ		98,8%	90,6%	83,5%	84,7%
Регион pol, REGA/Comet	98,8%		91,8%	84,7%	84,7%
Регион env, филогенетический анализ	90,6%	91,8%		90,6%	90,6%
Регион env, geno2pheno	83,5%	84,7%	90,6%		91,8%
Регион env, Comet	84,7%	84,7%	90,6%	91,8%	

которой представлен фрагментом субтипа А1, регион pol — фрагментами субтипов А и С. При проведении филогенетического анализа по обоим регионам эти образцы кластеризовались в одной группе с субтипом 02_AG. Однако результат программы geno2pheno, выдаваемый в виде «А or AG» не позволяет разделить чистый субтип и рекомбинант.

Кроме того, по результатам филогенетического анализа, в нашей выборке были обнаружены образцы, структура генома которых не позволила отнести их к известным рекомбинантам. Это образцы вариантов polBenvA и polABenvA. Для определения являются ли эти дискорданты вариантом рекомбинанта 03_AB с дополнительным фрагментом субтипа А в регионе env или это ошибки анализа, требуются дополнительные исследования.

Остальные дискордантные результаты явно представляют собой ошибки работы алгоритмов — это образцы, по филогенетическому анализу обоих регионов принадлежащие к субтипу А1, однако субтипированные в программах как С, D или CRF_01, 06, 35.

Уровень сходимости различных методов субтипирования по разным регионам представлен в таблице 2.

Основными недостатками автоматического субтипирования только по региону env в программах geno2pheno_[coreceptor] и Comet, является низкая чувствительность к рекомбинантам (в нашем случае к 02_AG и 03_AB). Кроме того, выявляются ошибки работы самих программ.

В целом, на исследуемой выборке наилучший уровень совпадения результатов субтипирования по разным регионам составил 91,8% при использовании филогенетического анализа нуклеотидных последовательностей. Самый низкий уровень совпадения был 83,5% при автоматическом субтипировании с помощью алгоритма geno2pheno региона env.

Необходимо отметить, что точность определения субтипа повышается пропорционально увеличению протяженности анализируемого фрагмента генома. Поэтому можно ожидать, что в случае наличия более протяженной нуклеотидной последовательности фрагмента региона env будет получена более высокая точность субтипирования.

ЛИТЕРАТУРА

1. *Global report: UNAIDS Report on global AIDS epidemic 2013 / WHO Library Cataloguing-in-Publication Data.*— Geneva: UNAIDS, 2013.— 198 p.
2. *Покровский В.В., Ладная Н.Н., Соколова Е.В., Буравцова Е.В.* ВИЧ-инфекция: Информационный бюллетень № 38.— М.: Федеральный научно-методический центр по профилактике и борьбе со СПИДом, 2013.— 53 с.— URL:http://hivruussia.ru/files/bul_38.pdf (дата обращения 07.07.2014) г.
3. *HIV DATABASES.*— URL: www.hiv.lanl.gov (дата обращения 5.06.2014).
4. *Hemelaar J.* Implications of HIV diversity for the HIV-1 pandemic // *Journal of Infections.*— 2013.— Vol. 66.— P. 391–400.
5. *Bobkova M.* Current status of HIV-1 diversity and drug resistance monitoring in the former USSR// *AIDS Rev.*— 2013.— Vol. 15.— P. 204–212.
6. *Santoro M.M., Perno C.F.* HIV-1 genetic variability and clinical implication // *ISRN Microbiology.*— 2013.— 2013.— P. 481314, doi: 10.1155/2013/481314.
7. *de Oliveira T., Deforche K., Cassol S., Salminen M., Paraskevis D., Seebregts C., Snoeck J., van Rensburg E.J., Wensing A.M., van de Vijver D.A., Boucher C.A., Camacho R., Vandamme A.M.* An automated genotyping system for analysis of HIV-1 and other microbial sequences // *Bioinformatics.*— 2005.— Vol. 21 (19).— p. 3797–3800.
8. *Lengauer T.* Bioinformatical assistance of selecting anti-HIV therapies: where do we stand? // *Intervirology.*— 2012.— Vol. 55.— P. 108–112.
9. *Tamura K., Stecher G., Peterson D., Filipski A., Kumar S.* MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0 // *Molecular Biology and Evolution.*— 2013.— Vol. 30.— P. 2725–2729.
10. *Lengauer T., Sander O., Sierra S., Thielen A., Kaiser R.* Bioinformatics prediction of HIV coreceptor usage // *Nat Biotechnol.*— 2007.— Vol. 25.— P. 1407–1410.
11. *Liitsola K., Holm K., Bobkov A., Pokrovsky V., Smolskaya T., Leinikki P., Osmanov S., Salminen M.* An AB recombinant and its parental HIV type 1 strains in the area of the former Soviet Union: low requirements for sequence identity in recombination. UNAIDS Virus Isolation Network // *AIDS Res. Hum. Retroviruses.*— 2000.— Vol. 16 (11).— P. 1047–1053.
12. *Yebra G., Mulder M., Martin L., Perez-Cachafeiro S., Rodriguez C., Labarga P.* Sensitivity of seven HIV subtyping tools differs among subtypes/recombinants in the Spanish cohort of naïve HIV-infected patients (CoRIS)// *Antiviral Res.*— 2011.— Vol. 89.— P. 19–25.
13. *Snoeck J.I., Kantor R., Shafer R.W., Van Laethem K., Deforche K., Carvalho A.P.* Discordances between interpretation algorithms for genotypic resistance to protease and reverse transcriptase inhibitors of human immunodeficiency virus are subtype dependent // *Antimicrob Agents Chemother.*— 2006.— Vol. 50.— P. 694–701.
14. *Struck D., Lawyer G., Ternes AM., Schmit JC., Perez Bercoff D.* COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification// *Nucleic Acids Research* 2014. doi 10.1093/nar/gku739.

Статья поступила 21.01.2015 г.

Контактная информация: Дмитриюкова Марина Юрьевна, e-mail: mdmitryukova@cmd.su

Коллектив авторов:

Дмитрюкова Марина Юрьевна — к.б.н., н.с. Центрального НИИ эпидемиологии Роспотребнадзора, 111123, Москва, ул. Новогиреевская, д. 3а, +7(495) 974-96-46, доб. 2601, e-mail: mdmitryukova@cmd.su.

Киреев Дмитрий Евгеньевич — к.б.н., н.с. Центрального НИИ эпидемиологии Роспотребнадзора, 111123, 111123, Москва, ул. Новогиреевская, д. 3а, +7(495) 974-96-46, доб. 2227, e-mail: dmitry.kireev@rcg.ru

Лопатухин Алексей Эдуардович — м.н.с. Центрального НИИ эпидемиологии Роспотребнадзора, 111123, Москва, ул. Новогиреевская, д. 3а, +7 (495) 974-96-46, доб. 2265, e-mail: loratukhin@rcg.ru.

Лаповок Илья Андреевич — к.б.н., н.с. Центрального НИИ эпидемиологии Роспотребнадзора, 111123, Москва, ул. Новогиреевская, д. 3а, +7 (495) 974-96-46, доб. 1121, e-mail: i_larovok@mail.ru.

Шипулин Герман Александрович — к.м.н., заведующий отделом молекулярной диагностики и эпидемиологии Центрального НИИ эпидемиологии Роспотребнадзора, 111123, Москва, ул. Новогиреевская, д. 3а, e-mail: gertman@rcg.ru.

ПОДПИСНЫЕ ИНДЕКСЫ:
ОБЪЕДИНЕННЫЙ КАТАЛОГ «ПРЕССА РОССИИ» — 42178
КАТАЛОГЕ АГЕНТСТВА «РОСПЕЧАТЬ» — 57990