Pcr.news: «Атака ложных родственников» может взломать генетические сервисы

Сайты, куда пользователи сами загружают генетическую информацию, например, для генеалогических исследований, могут быть уязвимы для хакеров. Злоумышленник, загрузивший на такой сайт чужие данные из открытого доступа или искусственно сконструированные «участки генома», получит информацию о генетических особенностях «родственников», в том числе связанных со здоровьем.

Елена Клещенко

опубликовано на сайте Pcr.news

Генетические тесты «напрямую к потребителю» (direct-to-consumer, DTC) набирают популярность по мере того, как дешевеют технологии исследования ДНК. В базах пяти наиболее крупных компаний уже сейчас присутствуют данные более 26 млн человек. Среди услуг, которые предоставляют такие компании, одна из самых популярных — поиск родственников; при этом часто есть возможность получить не только полные имена потенциальной родни, но и контактные данные. Во многих случаях пользователь не обязан присылать образец своего биоматериала, а может самостоятельно загрузить на сайт свои генетические данные, например, полученные в другой компании.

Майкл Эдж и Грэм Куп из Калифорнийского университета в Дэвисе описывают три способа, которые теоретически позволяют человеку с дурными намерениями получить доступ к данным других людей, в том числе к аллельным вариантам, ассоциированным с болезнями. Все три атаки, описанные в их статье, опубликованной в eLife, может осуществить любой, кто обладает знаниями в области генетики и биоинформатики.

Во всех трех случаях авторы используют понятие сегментов, идентичных по состоянию (identical-by-state, IBS), то есть таких сегментов, последовательности которых одинаковы или сходны у двух людей. IBS могут быть (но могут и не быть) идентичными по происхождению (identical-by-descent, IBD), то есть унаследованными от общего предка. Алгоритмы генеалогических сервисов ищут такие участки в геномах пользователей, по результатам поиска делают предварительные выводы о родстве и предоставляет клиентам соответствующую информацию, часто и о том, какие конкретно участки геномов совпали. Эдж и Куп показали, как этим может воспользоваться злоумышленник.

Первый способ — «замощение», или IBS tiling: на сайт загружаются множество реальных генотипов, взятых из открытых источниках. Для каждого из генотипов сервис выдает соответствия, и, сопоставив результаты по всем генотипам, можно собрать по кусочкам значительную часть генома конкретного лица из базы данных.

Второй способ, «зонд» (IBS probing), назван по аналогии с методом ДНК-гибридизации — использованием меченого фрагмента ДНК, чтобы найти в образце комплементарный участок. В этом случае загружают искусственную последовательность ДНК: она содержит интересующий хакера сайт (например, аллель, ассоциированный с болезнью Альцгеймера), окруженный специально сгенерированными последовательностями, в которых заведомо нет IBS-сегментов с реальными геномами. Таким образом, все лица, найденные поиском по базе данных, имеют этот аллель, и, следовательно, предрасположенность к заболеванию.

Третий способ, «приманка» (IBS baiting), использует полностью поддельные последовательности. Цель их — обмануть алгоритмы, которые ищут IBS сегменты по отсутствию несовместимых гомозиготных сайтов (то есть таких, которые у одного из сравниваемых лиц гомозиготны по одному аллелю, у другого — по другому); гетерозиготные сайты считаются совпадающими независимо то того, какой аллель в какой хромосоме. В этом случае загружаются два генотипа, гетерозиготные по всем сайтам, кроме интересующего; в одном из генотипов он гомозиготен по одному аллелю, в другом — по другому. Очевидно, что сравнение с этими двумя генотипами позволяет узнать статус данного сайта для любого генома. По подсчетам Эджа и Купа, из некоторых баз таким образом можно выкачать большую часть информации, загрузив всего около ста генотипов. Сходный метод описали почти одновременно с ними другие авторы: идея витает в воздухе.

В качестве проверки концепции авторы работы использовали «приманку» для идентификации конкретных SNP на сервисе GEDmatch (том самом, чьи данные помогли арестовать Убийцу из Золотого Штата). Это делалось в исследовательском режиме, чтобы загруженная «приманка» не взаимодействовала с данными других пользователей. Эксперимент удался.

«Люди отдают больше информации, чем думают», когда загружают свои генетические данные на общедоступные сайты, говорит Грэм Куп. Разумеется, соавторы проинформировали GEDmatch и другие подобные ресурсы о полученных результатах до публикации и предложили способы борьбы с возможными хакерскими атаками.

Источники

Michael D. Edge, Graham Coop. // Attacks on genetic privacy via uploads to genealogical databases. // eLife 2020; 9: e51810; DOI: 10.7554/eLife.51810

Цитата по пресс-релизу