

МНОГОМЕРНЫЙ АНАЛИЗ МИКРОЧИПОВЫХ ДАННЫХ ДЛЯ ВЫЯВЛЕНИЯ ПОТЕНЦИАЛЬНЫХ МАРКЕРОВ БОЛЕЗНИ ХАНТИНГТОНА

И.И. Шайдуллин

Казанский государственный университет

ildarshidullin@gmail.com

А.В. Катохин

Институт цитологии и генетики СО РАН

katokhin@bionet.nsc.ru

В.М. Ефимов

Институт цитологии и генетики СО РАН

efimov@bionet.nsc.ru

Применен новый способ многомерного анализа микрочиповых данных. Выявлено два независимых фактора, сочетание которых приводит к болезни Хантингтона. По первому фактору разделяются больные и здоровые и он предположительно связан с накоплением CAG повторов. Вторым фактором, дающим разницу между двумя группами (больные+здоровые, субнормальные+предрасположенные), по-видимому, являются изменения в иммунной системе человека. Для обоих факторов выявлены и аннотированы гены-кандидаты.

Синдром Хантингтона – аутосомное доминантное генетическое заболевание нервной системы, характеризующееся постепенным развитием обычно в возрасте 35-50 лет и сочетанием прогрессирующего хореического гиперкинеза и психических расстройств. Заболевание вызывается увеличением копийности тринуклеотидного повтора, соответствующего кодону CAG в гене IT-15, и приводящего к синтезу дефектного белка с многократно умноженной аминокислотой – глутамином. В гене дикого, не мутантного, типа у разных людей присутствует разное количество CAG повторов, однако, когда число повторов превышает 36, часто развивается болезнь.

Исследователями была высказана гипотеза, что поскольку экспрессия хантингтина, включая его мутантные формы, наблюдается во всех тканях, следствием его соматического мутирования будет, в частности, нарушение дифференциальной экспрессии генов в клетках крови, которое может служить маркером болезни Хантингтона. Поэтому необходимы исследования, способствующие выявлению различий в экспрессии генов у здоровых людей и у людей с синдромом Хантингтона при использовании малоинвазивных методов сбора клинического материала.

Проанализированы данные о профилях экспрессии генов в образцах периферической крови, взятых у 12 пациентов с болезнью Хантингтона, 5 – с предсимптоматикой этой болезни (предрасположенных), 14 – здоровых [1]. В предыдущей работе из здоровых средствами неметрического многомерного шкалирования нами выделена группа субнормальных [2]. Данная группа оказалась близка к предрасположенным к развитию болезни, хотя имела нормальное количество CAG-повторов.

Были использованы две матрицы экспрессии генов – на основе платформ Codelink и Affymetrix. Для обеих матриц экспрессии построены матрицы евклидовых расстояний

между образцами и обе матрицы обработаны методом главных координат. Далее был проведен тест Мантеля и вычислен коэффициент корреляции между матрицами расстояний. Оба набора координат обработаны 2B-PLS алгоритмом для нахождения бикомпонент – пар линейных комбинаций, максимально соответствующих друг другу. Объединенная матрица расстояний вычислена через корень из сумм квадратов соответствующих расстояний в обеих матрицах. Далее было получено расположение на плоскости профилей экспрессии генов больных и здоровых.

Тест Мантеля показал хорошие результаты соответствия данных по двум платформам, $r=0,631$, $p=1.0 \cdot 10^{-8}$ при 10^8 пермутаций.

Выделено две пары бикомпонент, на которые приходится 53,9% и 20% ковариации, соответственно.

Можно предположить, что в развитии синдрома Хантингтона участвует как минимум два фактора. И болезнь развивается при сочетании этих факторов.

Отображение на плоскость профилей экспрессии совокупности генов по первым бикомпонентам каждого человека из исследуемой выборки показано на рис.1.

Как видно, данная пара бикомпонент разделяет людей на несколько групп.

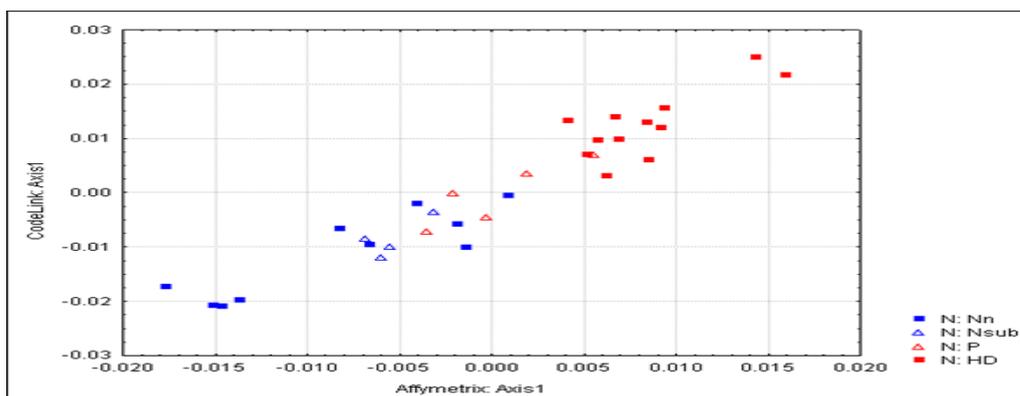


Рис. 1. Расположение образцов на плоскости первых бикомпонент.

Было предположено, что таким образом проявляется накопление CAG повторов. Группа здоровых людей, расположившаяся в нижней части рисунка 1, имеет минимальное количество повторов, группа больных людей, расположившаяся в верхней части рисунка – максимальное. Те, кто оказались между ними – имеют некие промежуточные значения количества CAG-повторов.

Значения экспрессии всех генов отображены на плоскость первых бикомпонент (рис. 2)

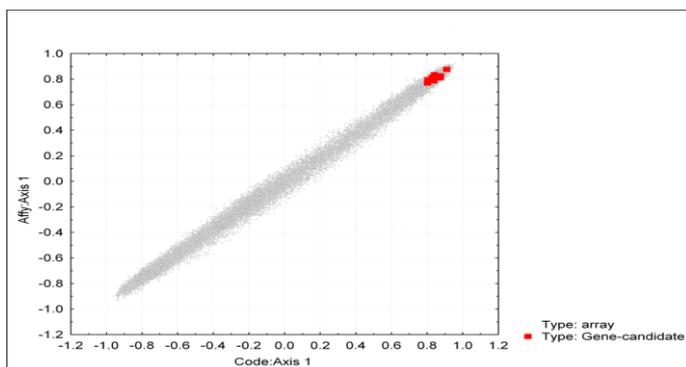


Рис. 2. Расположение профилей экспрессии генов на плоскости первых бикомпонент

Гены, расположенные в верхней правой части распределения, наиболее вовлечены в проявление болезни при анализе клеток периферической крови. Квадратами обозначены 12 генов-кандидатов, выявленных в работе [1] как гены-маркеры болезни Хантингтона.

Наши исследования, проведенные с использованием других алгоритмов анализа, подтверждают эти данные.

Из всех генов нами рассмотрено 105 генов, наиболее вовлеченных в проявление болезни. Показано, что эти гены наиболее часто указываются, как связанные с синдромом Альцгеймера, умственной отсталостью, различными нейродегенеративными синдромами, параличом и раком.

Отображение на плоскость профилей экспрессии совокупности генов по вторым бикомпонентам каждого человека из исследуемой выборки показано на рис.3.

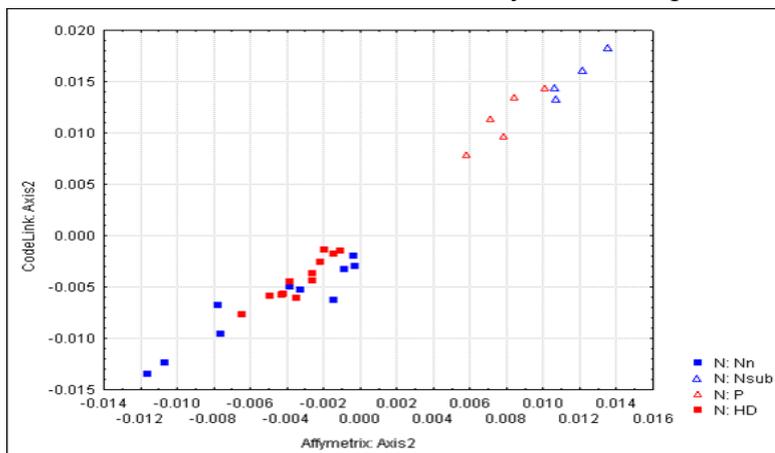


Рис. 3. Расположение образцов на плоскости вторых бикомпонент (абсцисса – платформа Affymetrix, ордината – платформа CodeLink)

Как видно из рисунка 3, все образцы разделились на две группы. В одну группу вошли больные и здоровые, в другую предрасположенные и субнормальные. Выделилась группа здоровых людей с нормальным количеством CAG повторов (субнормальные), которые оказались близки к предрасположенным к развитию болезни, т.е. между ними есть что-то общее. Было предположено, что есть некий механизм, который защищает от развития синдрома Хантингтона даже при накоплении критического числа CAG повторов, которое в норме дает проявление болезни. Причем данный механизм работает и у тех, кто имеет нормальное число повторяющихся кодонов. И субнормальная группа людей защищена дважды – у них нормальное количество CAG повторов, и работает пока неизвестный

механизм защиты. У больных и у группы, выступившей контролем, этот механизм не задействован, поэтому они оказались рядом.

На рисунке 4 показано расположение профилей экспрессии генов по вторым бикомпонентам. Гены, расположенные в правой части распределения, наиболее вовлечены в проявление пока неизвестного защитного механизма.

Синим треугольником показан ген *vezf1*, чья экспрессия наиболее вовлечена в проявление защитного механизма. У людей, оказавшихся в правой части рисунка 4, его экспрессия наиболее выражена.

Также были проанализированы 65 генов, чей уровень экспрессии наиболее близок по второй бикомпоненте к гену *vezf1* и, вероятно, также играет свою роль в защитном механизме.

Оказалось, что эти гены чаще всего упоминаются в статьях, посвященных исследованию иммунодефицита, некрозиса, различных инфекций, также часто встречался термин «лимфоцитарный».

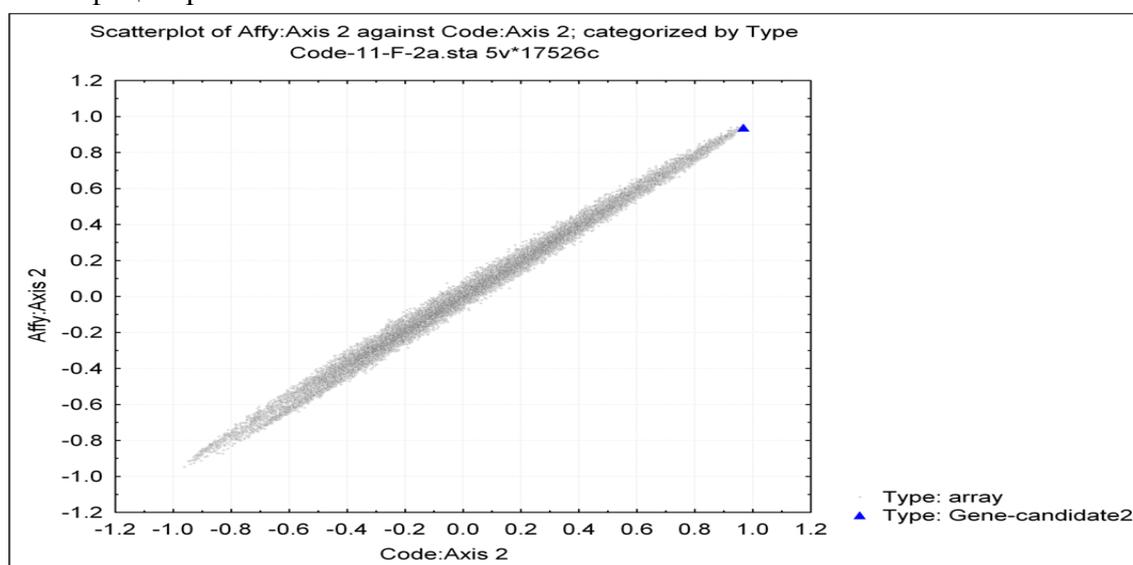


Рис. 4. Расположение профилей экспрессии генов на плоскости вторых бикомпонент

Мы предположили, что вторым фактором, дающим разницу между двумя группами (больные+здоровые, субнормальные+предрасположенные) являются изменения в иммунной системе человека. И именно этим может объясняться тот факт, что у некоторых людей, даже при наличии большого числа САГ повторов, не развивается болезнь.

У некоторых здоровых людей также есть эти изменения в иммунной системе и поэтому они оказываются близки с предрасположенными. Зная об этой особенности, представляется возможным разработать ранние и малоинвазивные методы диагностирования синдрома Хантингтона. А также представляется возможным разработать методы лечения этой болезни – если удастся внести некие изменения в иммунную систему – человек будет предохранен от развития болезни, даже имея наследственную предрасположенность к этому.

Для гена *vezf1* составлена цепь взаимодействий с другими генами из списка, чья экспрессия обеспечивает защиту от синдрома Хантингтона.

Получилась следующая цепь генов: *vezf1* связан через предсказанное белковое взаимодействие с *znf24*, для которого также предсказано взаимодействие с *actg1*. А этот ген в

свою очередь взаимодействует по тому же принципу с *crebbp*. Для этого гена указано физическое и химическое взаимодействие с *rela*. Такое же взаимодействие у этого гена с *ep300*.

ЛИТЕРАТУРА

[1]. Borovecki F, Lovrecic L, Zhou J, Jeong H, Then F, Rosas HD, Hersch SM, Hogarth P, Bouzou D, Jensen RV, Krainc D. Genome-wide expression profiling of human blood reveals biomarkers for Huntington's disease. Proc Natl Acad Sci USA, 2005. V. 102. P. 11023-11028.

[2]. Ефимов В.М. Катохин А.В. Применение неметрического многомерного шкалирования для мультиплатформенной обработки микрочиповых экспрессионных данных. Вестник ВОГиС. 2009. Т. 13, №1, С. 102–108.