Вопросы медицинской химии, 2002, том 48 №, с.353-356

ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

УДК 612.015(088.9)
©Лохов.

ЧИСЛЕННО-ЭВРИСТИЧЕСКИЙ ПОДХОД В АНАЛИЗЕ ПРОТЕОМНЫХ КАРТ

П.Г. Лохов.

ГУ НИИ Биомедицинской химии им. В.Н. Ореховича РАМН
119992, Москва, ул. Погодинская, д.10, факс (095)2450857

В данной работе исследованы возможности численно-эвристического подхода в анализе двумерных протеомных карт, реализованного на основе обработки нейросетью двумерных белковых электрофореграмм. На примере алкогольного синдрома плода оценена относительная эффективность подхода в диагностике заболеваний.

Ключевые слова: протеомика, экспертная система, двумерный электрофорез, диагностика, алкогольный синдром плода, нейросеть.

ВВЕДЕНИЕ. Интенсивно развивающаяся программа по инвентаризации белков - протеомика несет в себе высокую функциональность, связанную с тем, что структурные и метаболические изменения в биоматериале отражаются на его белковом составе. Это позволяет использовать ее для изучения структурно-функциональной организации клеток, генетических и фармакологических отклонений, действия лекарств и механизмов их токсичности, диагностики и мониторинга заболеваний, поиска новых биологически активных веществ. Благодаря своей функциональности, протеомика вызвала огромный интерес у ученых всего мира, что привело к созданию множества баз данных протеомных карт, полученных двумерным электрофорезом (SWISS-2DPAGE, SIENA-2DPAGE и др.). Однако помимо широких возможностей двумерный электрофорез имеет и существенные ограничения: в частности, трудности в разделении гидрофобных белков (например, мембранных белков) и идентификации минорных белков [1], которые в основном являются функционально важными ферментами. Эти ограничения связаны с лимитированной загрузочной емкостью полиакриламидных гелей, плохой растворимостью гидрофобных белков в стандартных растворах [2,3] и недостаточно высокой чувствительностью метода. Так, при использовании радиоактивной метки для визуализации белков или при сужении диапазона pH при изоэлектрофокусировании обнаруживается в несколько раз больше белков, чем при стандартных условиях получения протеомных карт [4]. Все эти недостатки в совокупности с высокой вариабельностью двумерных электрофореграмм (ЭФГ) и низким процентом идентифицируемых на них белков существенно затрудняют прикладное использование протеомных карт. Нами рассмотрен один из возможных вариантов решения вышеперечисленных проблем, основанный на обнаружении

мотивов в изменении белкового состава протеомных карт при различных процессах в исследуемом биоматериале (патологический процесс, мутация, лекарственная устойчивость и т.д.). Отсутствие при данном подходе необходимости знания функций белков, биомаркеров, а также визуализации всех белков, входящих в биоматериал, может кардинально упростить анализ протеомных карт и прикладное применение протеомики.

МЕТОДИКА. *Двумерные электрофореграммы*. В работе использовали 8 двумерных ЭФГ плазмы крови (изоэлектрофокусировка в диапазоне pH 4,5-8,5, SDS-PAGE в диапазоне Mr 10-250 кДа) здоровых девочек и мальчиков 9-11 лет и больных алкогольным синдромом плода (АСП) [5] из базы данных Fliker/NCI (www.lecb.ncifcrf.gov/flicker). На рисунке 1 указан участок двумерной ЭФГ, использованный для анализа, и денсиграмма этого участка, на которой представлены профили таких белков как фибриноген (внешняя гамма-цепь), альфа-цепь IgA, альфа-1-антитрипсин, бета-цепь гаптоглобулина, апо-А-1.

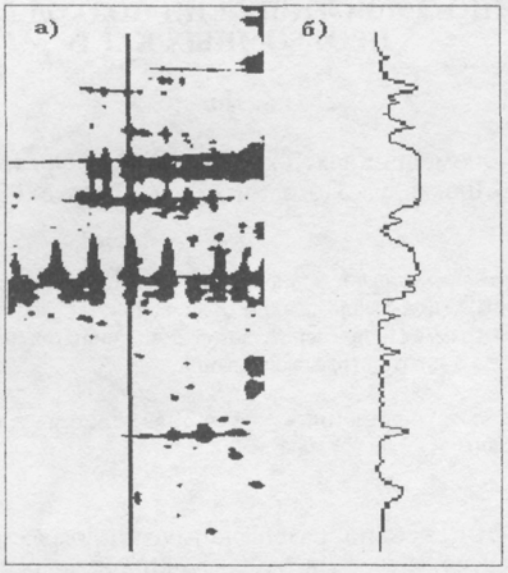

Рисунок 1.
Средняя часть двумерной ЭФГ (а) плазмы крови 10-ти летнего мальчика. Линией указан участок, использованный для снятия денсиграммы (б).

Анализ электорофореграмм. Для анализа ЭФГ использовали экспертную систему Proteometrics (ГУ НИИ биомедицинской химии им. В.Н. Ореховича РАМН), способную по денсиграммам выявлять мотивы в изменении двумерных и одномерных ЭФГ. Ядро программы построено на архитектуре нейросети (НС), вариант Time Lagged Recurrent Network [6]. Денсиграммы, снятые с ЭФГ перед введением в НС, нормализовали (максимальное значение - 1, минимальное - 0). Обучение экспертной системы распознавать мотивы в изменении денсиграмм проводили два раза (по 50 тыс. циклов) на денсиграммах, соответствующих здоровым и больным мальчикам. Тестирование системы проводили на денсиграммах как здоровых и больных мальчиков, так и девочек. Критерием распознавания мотива при обучении и тестировании НС являлось среднеквадратическое отклонение (СКО) между предсказанными и реальными значениями точек профилей электрофоретических денсиграмм.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ. Своевременное медикаментозное лечение больных с АСП (встречаемость 2 на 1000 новорожденных) затруднено в связи с отсутствием специфического теста диагностики этого синдрома. Поэтому неудивительно, что была предпринята попытка диагностировать эту патологию методами протеомики [5]. В результате обнаружено 8 белков, которые

классифицировали как потенциальные биомаркеры АСП с достоверностью $p<0,01$, и 21 белок с достоверностью $p<0,02$ (t-тест Стьюдента). В связи с чем для нас представляло особый интерес провести численно-эвристический анализ этих же протеомных карт с целью сравнения его возможностей с диагностикой по биомаркерам. Результаты обучения НС в распознавании денсиграмм, соответствующих норме и патологии мальчиков, приведены в табл. 1. Результаты тестирования НС представлены в табл. 2, из которых следует, что обученная на ЭФГ плазмы здоровых мальчиков НС хорошо распознает как участвовавшую в ее обучении ЭФГ, так и не участвовавшую (контрольную) ЭФГ (табл.2, опыты 1, 2, 3). Распознавание ЭФГ плазмы больных мальчиков на 2 порядка ниже (табл.2, опыты 4, 5, 6). Причем СКО для них практически идентичное. Таким образом, НС выделила мотивы как нормы, так и патологии (АСП), с достоверностью отличия $p<0,000005$ (t-тест Стьюдента). Что касается ЭФГ девочек, то АСП не распознан, т.к. ЭФГ девочек в обучении НС не участвовали, однако обнаруженный в них мотив (практически идентичное СКО для опытов 7 и 8) четко дифференцируется от мотивов ЭФГ мальчиков.

Таблица 1. Результаты обучения НС.

Выриант обучения НС	Кол-во денсиграмм	СКО
на ЭФГ плазмы здоровых мальчиков	2	$5,835 \cdot 10^{-8}$
на ЭФГ плазмы больных мальчиков	2	$5,649 \cdot 10^{-8}$

Примечание: СКО- среднеквадратичное отклонение.

Таблица 2. Результаты тестирования НС.

ЭФГ плазмы	№ опыта	СКО
Для обучения НС на ЭФГ плазмы здоровых мальчиков		
здоровых мальчиков	1	$\mathbf{1,122 \cdot 10^{-7}}$
	2	$\mathbf{1,122 \cdot 10^{-7}}$
	3	$1,136 \cdot 10^{-7}$
больных мальчиков	4	$1,063 \cdot 10^{-5}$
	5	$1,065 \cdot 10^{-5}$
	6	$1,061 \cdot 10^{-5}$
здоровых девочек	7	$9,349 \cdot 10^{-6}$
больных девочек	8	$9,354 \cdot 10^{-6}$
Для обучения НС на ЭФГ плазмы больных мальчиков		
здоровых мальчиков	10	$2,838 \cdot 10^{-6}$
	11	$2,840 \cdot 10^{-6}$
больных мальчиков	12	$\mathbf{9,157 \cdot 10^{-8}}$
	13	$\mathbf{7,572 \cdot 10^{-8}}$
	14	$91,64 \cdot 10^{-8}$
здоровых девочек	15	$1,422 \cdot 10^{-6}$
больных девочек	16	$1,427 \cdot 10^{-6}$

Примечание: Выделенны СКО для ЭФГ, участвовавших в обучении НС.

Несколько иная ситуация в случае обучения НС на денсиграммах ЭФГ плазмы больных мальчиков. АСП у мальчиков распознан довольно хорошо, но с разным СКО (табл. 2, опыт 12, 13), что указывает на не идентичность мотивов. Скорее всего, это связано с индивидуальными отличиями течения болезни у этих пациентов (степень тяжести АСП, сопутствующие заболевания и т.д.), которые отразились на ЭФГ. Не идентичность мотивов, участвовавших в обучении НС,

привела к существенной ошибке в распознавании контрольной ЭФГ (табл. 2, опыт 14). Это указывает на то, что в случае создания диагностических систем следует избегать неоднородности мотивов, используемых для обучения НС. В остальном НС повторяет свои уже продемонстрированные свойства. Достоверно определен мотив нормы мальчиков, который дифференцируется от мотива, определяемого женским полом, и не определяет состояние нормы и патологии у девочек.

Суммировав вышесказанное, можно сделать следующие выводы:

- достоверность постановки диагноза по протеомным картам численно-эвристическим методом на несколько порядков выше, чем при использовании биомаркеров болезни ($p<0,000005$ против $p<0,01$);
- численно-эвристическим методом можно определить чистоту мотива в изменении белкового состава биоматериала (очень важно для анализа протеомных карт при инвентаризации белков при норме и различных патологиях);
- для создания диагностической системы определенного состояния (патологический процесс, мутация, лекарственная устойчивость и т.д.) необходимо всего две протеомные карты дающие идентичные СКО, в отличие от диагностики по биомаркерам, требующим выборки для статистической обработки;
- численно-эвристический метод дает возможность диагностировать неизвестные состояния, когда СКО отличается от контроля и опыта (нормы или патологии) и при этом идентично для двух протеомных карт.

ЛИТЕРАТУРА

1 *Wilkins M.R., Gasteiger E., Sanchez J.C., Bairoch A., Hochstrasser D.F.* (1998) Electrophoresis, **19**,1501-1505.

2 *Adessi C., Miege C., Albrieux C., Rabilloud T.* (1997) Electrophoresis, **18**, 127-135.

3 *Chevallet M., Santoni V., Poinas A., Rouqui D., Fuchs A., Kieffer S., Rossignol M., Lunardi J., Garin J., Rabilloud T.* (1998) Electrophoresis, **19**, 1901-1909.

4 *Westbrook J.A., Yan J.X., Wait R., Welson S.Y., Dunn M.J.* (2001) Electrophoresis, **22**, 2865-2871.

5 *Robinson M.K., Myrick J.E., Henderson L.O., Coles C.D., Powell M.K., Orr G.A., Lemkin P.F.* (1995) Electrophoresis, **16**, 1176-83.

6 *deVries B., Principe J.* (1992) Neural Networks, **5**, 565-576.

Поступила 09.09.01.

NUMERICAL-HEURISTIC APPROACH FOR ANALYSIS OF PROTEOME MAPS

P.G. Lokhov

Orekhovich Institute of Biomedical Chemistry RAMS
Pogodinskaya st., 10, Moscow, 119121, Russia; fax (095)2450857

The possibility of numerical-heuristic approach has been examined for proteome map analysis that realized by neural network processing of 2D protein electrophoregramms. The fetal alcohol syndrome was used to illustrate that this approach islatively effective for diagnostics of diseases.

Keywords: proteomics, expert system, two-dimensional electrophoresis, diagnostics, fetal alcohol syndrome, neural network.