

Предсказание генов маркеров агрессивности глиомы низкой степени злокачественности по экспрессионным данным TCGA

Иванисенко Н.В.^{*1,2}, Губанова Н.В.^{1,2}, Колчанов Н.А.¹,

Иванисенко В.А.¹

¹Институт цитологии и генетики Сибирского отделения РАН, Новосибирск, 630090, Россия

²Новосибирский Государственный Университет, Новосибирск, 630090, Россия

Аннотация. Глиома низкой степени злокачественности (ГНСЗ) представляет собой группу опухолей ЦНС. В отличие от других злокачественных субтипов, таких как мультиформная глиобластома, глиомы низкой степени злокачественности имеют более низкую смертность. Однако после первичного лечения ГНСЗ наблюдается около 40% случаев повторного образования опухолей. В данной работе с использованием RNA-seq данных из базы данных The Cancer Genome Atlas (TCGA) был проведен поиск генов, ответственных за агрессивный фенотип опухолей. Было обнаружено 85 генов, нарушение в уровне экспрессии которых ассоциировано с появлением новых опухолей после первичного лечения.

Ключевые слова: глиома низкой степени злокачественности (ГНСЗ), глиома, RNA-seq, агрессивность, TCGA.

ВВЕДЕНИЕ

Глиома низкой степени злокачественности (ГНСЗ) является инфильтративной опухолью головного мозга, характеризующейся бессимптомным развитием, часто охватывающим функционально критические участки мозга. В настоящее время золотым стандартом постановки диагноза опухолевых заболеваний является гистологическая классификация, приведенная в табл. 1. Диагностический подход, основанный исключительно на гистологических критериях, имеет ряд недостатков, одним из наиболее важных которых является отсутствие воспроизводимости. Например, диагностические разногласия между четырьмя экспертами-невропатологами составили: 43% при астроцитомах II степени, 64% при анапластических астроцитомах и 38% при глиобластомах [1]. Недостаток воспроизводимости гистологической диагностики часто приводит к появлению противоречивой информации по эффективности лечения. В сравнении с другим злокачественным субтипом, мультиформной глиобластомой (степень IV согласно классификации всемирной организации здравоохранения (ВОЗ)), являющимся наиболее агрессивной формой, глиома низкой степени (степень II по классификации ВОЗ), имеющая три различных гистологических типа (диффузная астроцитома, олигоастроцитома и олигодендроглиома), характеризуется высокой средней выживаемостью. Известно, что сложность успешного лечения опухолевых заболеваний зависит от множества факторов. Одним из таких факторов является агрессивность опухоли, которая

* n.ivanisenko@gmail.com

определяется частотой возникновения новых случаев опухоли после лечения (в частности, метастаз, или возникновение опухолей более высокой степени злокачественности).

Таблица 1. Основные гистологические критерии классификации глиом ВОЗ за 2000 г.

Гистологический тип	Дифференциация	Клеточная плотность	Ядерная атипия	Митотическая активность	Некроз	Сосудистая пролиферация
Диффузная астроцитома, олигоастроцитома и олигодендроглиома степень II	Высокая степень дифференцировки	Умеренная	иногда	Нет или 1 митоз	нет	нет
Анапластическая астроцитома степень III	Анаплазия фокальная или генерализованная	Увеличена диффузно или локально	есть	есть	нет	нет
Глиобластома степень IV	слабая	высокая	выраженная	выраженная	есть	есть

В настоящее время накоплено огромное количество информации о молекулярных изменениях в злокачественных новообразованиях, представленной в базах данных и научной литературе, использование которой позволяет глубже понять механизмы злокачественной трансформации, а так же существенно улучшить постановку диагноза и прогноза, и, как следствие, оптимизировать процесс лечения. В настоящей работе с использованием данных из The Cancer Genome Atlas (TCGA) [2] был проведен поиск генов, потенциально ответственных за агрессивность опухолей ГНСЗ. Было обнаружено 90 генов, нарушение в уровне экспрессии которых статистически значимо коррелирует с существенным повышением частоты возникновения новых опухолей. Статистический анализ сверхпредставленности биологических процессов Gene Ontology показал связь с агрессивным фенотипом опухоли таких процессов, как «клеточная адгезия», «регуляция деполимеризации микротрубочек» и «развитие нервной системы». Согласно анализу литературных данных, большинство обнаруженных генов играют существенную роль в процессах миграции опухолевых клеток, многие из которых выделяются в качестве потенциальных мишеней для разработки лекарственных соединений.

МАТЕРИАЛЫ И МЕТОДЫ

Данные по уровню экспрессии генов (RNA-seq данные) в опухолевых тканях, образцы которых были получены от более 500 пациентов с ГНСЗ извлекались из базы данных The Cancer Genome Atlas [2].

Анализ сверхпредставленности биологических процессов Gene Ontology проводился с использованием модуля BiNGO в пакете Cytoscape [3, 4].

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Для идентификации потенциальных экспрессионных маркеров агрессивности ГНСЗ, были проанализированы RNA-seq данные опухолевых тканей пациентов с диагнозами, относящимися к ГНСЗ. На первом шаге анализа проводилась идентификация генов, имеющих несколько локальных максимумов в распределении уровня экспрессии по различным пациентам. Поиск локальных максимумов осуществлялся с помощью интерполяции распределения степенным полиномом. Минимальный уровень экспрессии между двумя пиками в таком распределении принимался равным пороговому значению, в соответствии с которым пациенты разбивались на две группы. Первая группа состояла из пациентов, у которых уровень экспрессии данного гена был ниже порогового значения, а вторая группа включала

пациентов с уровнем экспрессии, превышающим порог. Всего было выявлено более 500 генов, имеющих выраженное полимодальное распределение.

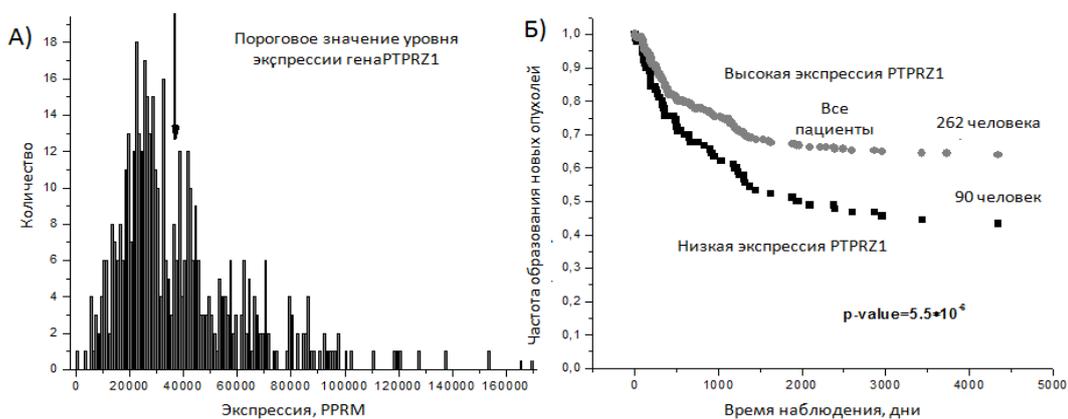


Рис. 1. Пример классификации пациентов по уровню экспрессии гена PTPRZ1. А) Распределение уровня экспрессии гена PTPRZ1. Пороговое значение уровня экспрессии показано стрелкой (>40000 PPRM). Б) Диаграмма Каплана-Майера, иллюстрирующая агрессивность опухолей у пациентов в группе с повышенным уровнем экспрессии гена PTPRZ1.

Сравнение уровня агрессивности опухолей для групп пациентов проводилось с помощью log-rank теста, который заключается в сравнении частот появления новых опухолей после первичного лечения, в зависимости от времени наблюдения за пациентами. Для визуализации разницы в частотах появления новой опухоли при сравнении групп пациентов использовались диаграммы Каплана-Майера. На рис. 1,А приведен пример распределения уровня экспрессии гена PTPRZ1 с указанием порогового значения (> 40000 PPRM), по которому проводилась классификация пациентов на две группы. Как видно из диаграммы Каплана-Майера, агрессивность опухолей у пациентов в группе с повышенным уровнем экспрессии гена PTPRZ1 существенно выше по сравнению со всей выборкой пациентов (рис. 1,Б).

Проведенный анализ позволил выявить 85 генов, нарушение в уровне экспрессии которых статистически значимо (статистическая значимость меньше $10E-5$) связано с возникновением новой опухоли после первичного лечения (табл. 2). Согласно таблице 2, можно отметить, что гены PTPRZ1, TIMP2 и NOTCH1 имели повышенные уровни экспрессии относительно соответствующих пороговых значений у более 20% всех пациентов. Среди генов, имеющих пониженный уровень экспрессии у большинства пациентов с агрессивным типом опухоли можно выделить ген PFN2. Полученные данные согласуются с результатами исследований других авторов. В частности, в работе [5] показано, что экспрессия гена PTPRZ1 имела корреляцию с частотой метастаз в астроцитомах и при раке желудка, а гены TIMP2, NOTCH2 играют критическую роль в развитии глиблостомы [6, 7]. Имеются данные, что белок PFN2 регулирует полимеризацию актина [8]. Как приведено выше, данный ген имел пониженный уровень экспрессии у пациентов с агрессивным фенотипом ГНСЗ, в связи с чем можно ожидать, что направленное увеличение его уровня экспрессии может рассматриваться как потенциальная стратегия действия новых лекарственных препаратов. Всего пониженным уровнем экспрессии характеризовались 39 генов из 70. Среди них гены SPARCL1, RASSF4 и PIK3R1 являются хорошо изученными генами-супрессорами опухолей [9–11]. Согласно работе [12] активация гена HSPA8 также приводила к более эффективному лечению опухоли. Следует отметить, что согласно базе данных ANDSystem, полученной с помощью автоматического анализа текстов научных публикаций [13], большинство выявленных в работе генов являются ассоциированными со многими типами злокачественных новообразований (данные не

показаны). Изучение роли вновь выявленных генов может быть предметом для проведения дальнейших исследований. Хорошее согласие полученных результатов с литературными данными позволяет сделать вывод о высокой предсказательной способности предложенного анализа.

Таблица 2. Список генов, изменение экспрессии которых ассоциировано с новыми случаями появления опухоли после первичного лечения согласно результатам анализа

Имя гена	% агрессивности ГНСЗ	N	Статистическая значимость	Имя гена	% агрессивности ГНСЗ	N	Статистическая значимость
PTRZ1	0.43333	90	5.59E-06	PLEC	0.47059	34	3.03E-10
TMP2	0.5	86	2.65E-05	FAM107A ↓	0.5	34	9.62E-11
NOTCH1	0.49398	83	1.76E-05	DDR1	0.45454	33	3.06E-11
PFN2 ↓	0.5	82	3.59E-06	LRP1	0.45454	33	9.15E-12
TMSL3	0.5	64	4.24E-07	CLIP2	0.42424	33	1.37E-11
SCG3 ↓	0.47541	61	1.64E-08	PDPN	0.34375	32	1.11E-16
SPP1	0.45902	61	3.67E-08	THRA ↓	0.4375	32	8.60E-13
ACTB	0.5	58	5.21E-08	NAP1L1 ↓	0.46875	32	1.59E-12
RTN1 ↓	0.44643	56	2.66E-09	NRXN1 ↓	0.46875	32	1.37E-11
SPARCL1 ↓	0.48214	56	1.10E-08	FXRD6 ↓	0.4375	32	2.34E-13
HSPB8 ↓	0.46429	56	3.69E-09	MAP4	0.48387	31	5.81E-11
GRIA1 ↓	0.49091	55	1.85E-08	PDGFRA	0.45161	31	3.71E-11
VIM	0.5	54	1.02E-08	SOX9	0.3871	31	5.77E-15
HSPA8 ↓	0.48	50	3.40E-09	RPL3 ↓	0.5	30	2.46E-11
PEA15 ↓	0.48936	47	6.78E-10	GPM6A ↓	0.43333	30	2.77E-13
PTN	0.45652	46	7.42E-11	NES	0.33333	30	1.11E-16
DBI	0.5	44	3.84E-09	PEA15	0.5	30	2.48E-10
HSP90B1	0.5	44	1.32E-09	RPLP1	0.5	30	5.58E-10
GPR56	0.41861	43	3.56E-11	ATP6V1G2 ↓	0.26667	30	0
CALCRL ↓	0.44186	43	1.62E-11	NTRK2 ↓	0.48276	29	2.39E-13
HMP19 ↓	0.41861	43	4.54E-12	PHYHIPL ↓	0.41379	29	5.11E-15
GABBR1 ↓	0.41861	43	5.51E-12	RASSF4 ↓	0.48276	29	1.61E-11
ATP1B2	0.42857	42	3.06E-11	NCAN	0.48276	29	3.06E-10
KCTD12	0.4878	41	2.88E-09	NRXN2	0.48276	29	1.93E-11
MSN	0.36585	41	1.02E-14	GLUD1 ↓	0.5	28	1.36E-11
C14orf132 ↓	0.45	40	1.23E-11	SPOCK2 ↓	0.42857	28	4.50E-14
SOX8 ↓	0.47368	38	2.33E-11	TRIL	0.5	28	2.05E-10
MAPT ↓	0.44737	38	6.73E-12	RND2 ↓	0.46429	28	7.44E-13
BCAN	0.47368	38	8.94E-10	PIK3R1 ↓	0.46429	28	1.99E-12
AEBP1	0.5	38	2.65E-10	RPS17	0.46429	28	2.58E-11
CST3	0.44737	38	2.37E-11	RPS16	0.48148	27	7.99E-11
KCNJ10 ↓	0.45946	37	2.10E-11	ITGB8	0.44444	27	6.23E-13
NDRG2 ↓	0.45946	37	1.08E-11	SDC3	0.48148	27	1.70E-11
TSC22D1 ↓	0.48649	37	3.99E-11	CLASP2 ↓	0.38461	26	1.22E-15
SERPINA3	0.43243	37	5.42E-12	FAIM2 ↓	0.42308	26	6.33E-14
LGALS3BP	0.48649	37	1.07E-10	LAMP2	0.5	26	1.67E-10
CH13L1	0.40541	37	4.88E-13	RPLP0	0.46154	26	1.08E-11
SULF2 ↓	0.5	36	8.71E-11	WSCD1	0.38461	26	3.33E-16
RUFY3 ↓	0.41667	36	4.31E-13	CRYAB ↓	0.46154	26	2.97E-13
GALNT13 ↓	0.38889	36	4.77E-15	IGFBP5	0.5	26	4.94E-12
CLU	0.45714	35	2.25E-11	ATCAY ↓	0.48	25	9.01E-13
APLP2	0.48571	35	2.72E-10	GJA1	0.44	25	1.14E-13
LHFPL3 ↓	0.5	34	1.31E-10				

Примечание: % агрессивности опухолей – процент пациентов в группе, имевших новые случаи возникновения опухолей после первичного лечения; N – число пациентов в данной группе; ↓ – гены, имеющие пониженный уровень экспрессии.

Анализ сверхпредставленности биологических процессов Gene Ontology позволил выявить биологические процессы, в которых участвует количество генов, из числа идентифицированных генов, превосходящее ожидаемое по случайным причинам. Среди таких биологических процессов можно выделить процесс «развитие нервной системы», в котором участвовали 25 из 85 генов (статистическая значимость 3.8E-10), 15 генов из 85 были вовлечены в процесс «клеточная адгезия» (статистическая

значимость $3.4E-6$), 3 гена из 85 участвовали в процессе «регуляция деполимеризации микротрубочек» (статистическая значимость $7.3E-5$).

Для выявления групп генов, потенциально обладающих общей регуляцией экспрессии, была построена сеть, в которой некий ген принимался связанным с другим геном, если экспрессия обоих генов превышала пороговое значение более чем у 70% пациентов (рис. 2). Выбор ограничения в 70% обеспечивал статистическую значимость взаимосвязи групп пациентов между собой, связанных с соответствующими парами генов. Статистическая значимость оценивалась с использованием точного теста Фишера.

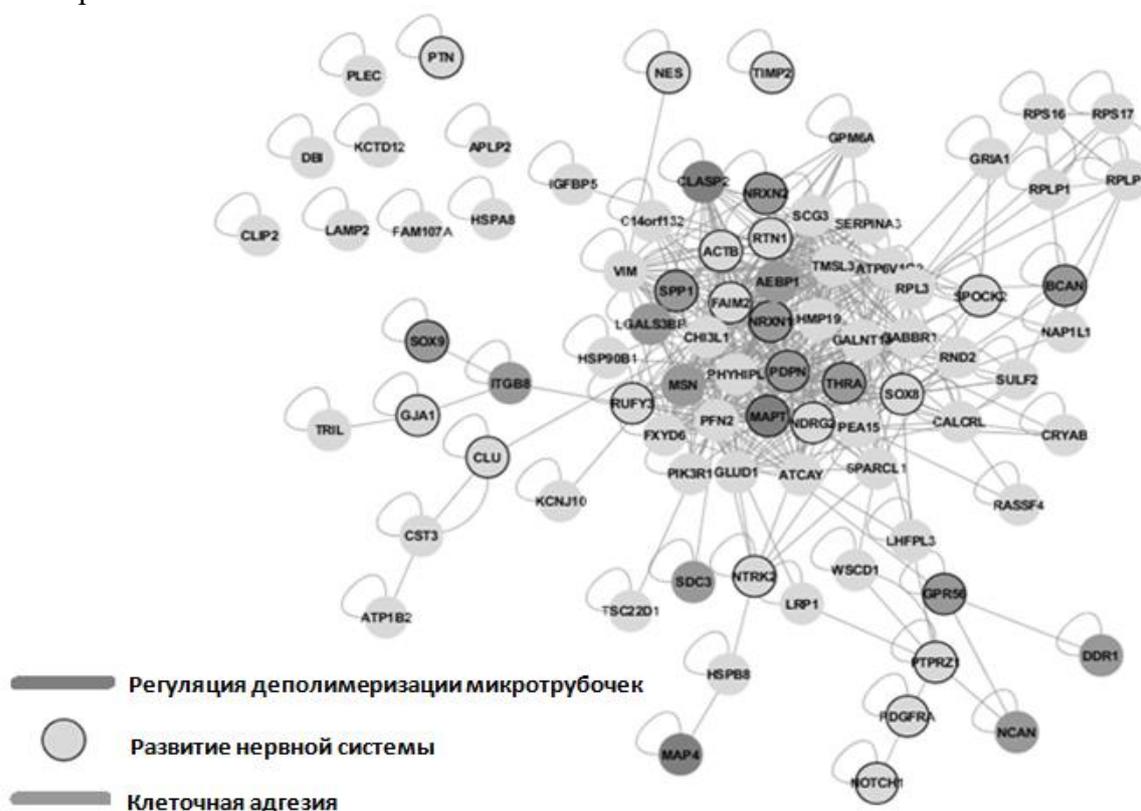


Рис. 2 Сеть взаимосвязей между генами, построенная на основе анализа уровня их экспрессии в тканях ГНСЗ. Тёмно-серым цветом показаны гены, участвующие в регуляции деполимеризации микротрубочек; светло-серым – гены, участвующие в клеточной адгезии; в черной рамке – гены, участвующие в развитии нервной системы. Сеть строилась в пакете Cytoscape [4].

Кластерный анализ построенной сети генов, связанных друг с другом по критерию сходства уровня экспрессии в тканях ГНСЗ, показал наличие группы из 37 генов, образующих ядро сети и потенциально имеющих общую регуляцию экспрессии. Согласно анализу сверхпредставленности процессов Gene Ontology, 8 генов из данного кластера были статистически значимо вовлечены в процесс «клеточная адгезия», а 2 гена (MAPT, CLASP2) оказались связанными с процессом «деполимеризации микротрубочек». Можно предположить, что высокая связанность данного кластера является следствием глобальных эпигенетических нарушений, которые приводят к комплексным изменениям в уровне экспрессии множества генов, ассоциированных с агрессивным фенотипом опухоли.

ЗАКЛЮЧЕНИЕ

С использованием данных RNA-seq TCGA был проведен поиск экспрессионных маркеров уровня агрессивности глиом низкой степени злокачественности. Для анализа были отобраны гены, форма распределения экспрессии которых при рассмотрении

выборки пациентов, имела несколько пиков (локальных максимумов), свидетельствующих о вариациях молекулярно-генетических процессов. Было сделано предположение, что наблюдаемая гетерогенность пациентов по уровню экспрессии определенных генов, может характеризовать отклонения по тяжести их заболевания. Всего было отобрано более 500 генов, обладающих полимодальной формой распределения уровня экспрессии. Оказалось, что около 20% из них статистически значимо ассоциированы с агрессивностью опухолей. Анализ Gene Ontology показал участие данных генов в процессах развития нервной системы, клеточной адгезии, а так же регуляции деполимеризации микротрубочек. Большинство из обнаруженных генов являются не только маркерами агрессивности развития опухолей, но и доказанными перспективными мишенями лечения различных злокачественных новообразований.

Вся работа по созданию и анализу модели осуществлялась при поддержке гранта РФ № 14-44-00011. Создание параллельных программ и суперкомпьютерные вычисления осуществлялись при поддержке проекта VI.61.1.2.

СПИСОК ЛИТЕРАТУРЫ

1. Mittler M., Walters B.C., Stopa E.G. Observer reliability in histological grading of astrocytoma stereotactic biopsies. *J. Neurosurg.* 1996. V. 85. P. 1091–1094.
2. McLendon R., Friedman A., Bigner D., Van Meir E.G., Brat D.J., Mastrogiannis G.M., Olson J.J., Mikkelsen T., Lehman N., Yung W.K.A et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature.* 2008. V. 455. P. 1061–1068.
3. Maere S., Heymans K., Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. *Bioinformatics.* 2005. V. 21. P. 3448–3449.
4. Smoot M.E., Ono K., Ruscheinski J., Wang P.L., Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics.* 2011. V. 27. № 3. P. 431–432.
5. Ma Y., Ye F., Xie X., Zhou C., Lu W. Significance of PTPRZ1 and CIN85 expression in cervical carcinoma. *Archives of gynecology and obstetrics.* 2011. V. 284. P. 699–704.
6. Lu K.V., Jong K.A., Rajasekaran A.K., Cloughesy T.F., Mischel P.S. Upregulation of tissue inhibitor of metalloproteinases (TIMP)-2 promotes matrix metalloproteinase (MMP)-2 activation and cell invasion in a human glioblastoma cell line. *Laboratory investigation.* 2003. V. 84. P. 8–20.
7. Fan X., Khaki L., Zhu T.S., Soules M.E., Talsma C.E., Gu N., Koh C., Zhang J., Li Y.M., Maciaczyk J. et al. NOTCH Pathway Blockade Depletes CD133-Positive Glioblastoma Cells and Inhibits Growth of Tumor Neurospheres and Xenografts. *Stem cells.* 2010. V. 28. № 1. P. 5–16.
8. Fan Y., Potdar A.A., Gong Y., Eswarappa S.M., Donnola S., Lathi J.D., Hambardzumyan D., Rich J.N., Fox P.L. Profilin-1 phosphorylation directs angiocrine expression and glioblastoma progression through HIF-1 α accumulation. *Nature cell biology.* 2014. V. 16. № 5. P. 445–456.
9. Esposito I., Kayed H., Keleg S., Giese T., Sage E.H., Schirmacher P., Friess H., Kleeff J. Tumor-suppressor function of SPARC-like protein 1/Hevin in pancreatic cancer. *Neoplasia.* 2007. V. 9. № 1. P. 8–17.
10. Donninger H., Vos M.D., Clark G.J. The RASSF1A tumor suppressor. *Journal of cell science.* 2007. V. 120. № 18. P. 3163–3172.
11. Taniguchi C.M., Winnay J., Kondo T., Bronson R.T., Guimaraes A.R., Alemán J.O., Luo J., Stephanopoulos G., Weissleder R., Cantley L.C., Kahn C.R. The

- phosphoinositide 3-kinase regulatory subunit p85 α can exert tumor suppressor properties through negative regulation of growth factor signaling. *Cancer research*. 2010. V. 70. № 13. P. 5305–5315.
12. José-Enériz E.S., Román-Gómez J., Cordeu L., Ballestar E., Gárate L., Andreu E.J., Isidro I., Guruceaga E., Jiménez-Velasco A., Heiniger A. BCR-ABL1-induced expression of HSPA8 promotes cell survival in chronic myeloid leukaemia. *British journal of haematology*. 2008. V. 142. № 4. P. 571–582.
 13. Demenkov P.S., Ivanisenko T.V., Kolchanov N.A., Ivanisenko V.A. ANDVisio: A new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem. *In Silico Biology*. 2012. V. 11. № 3. P. 149–161.

Материал поступил в редакцию 27.11.2014, опубликован 18.12.2014.