

Отбор биомаркеров злокачественных новообразований головного мозга и здоровых пациентов на основе профилей микроРНК в плазме крови по данным секвенирования нового поколения и машинного обучения

Гущина А. С.¹

Научный руководитель – канд. биол. наук, Пушкин А. А.^{2,3}

¹Университет ИТМО

²ООО “НПФ Синтол”

³ИАП РАН

nas.gushchina@mail.ru

Введение

Глиобластома является наиболее распространенной и агрессивной первичной опухолью как головного мозга, и может составлять от 40% до 50% от всех новообразований центральной нервной системы [1]. При этом, медиана выживаемости пациентов около 15 месяцев [2]. Также существенное распространение имеют и метастазы головного мозга, эффективность терапии которых в значительной степени определяется способностью надежно различать эти классы опухолей.

Традиционные методы диагностики и прогнозирования, основанные на клиническом осмотре, МРТ или гистологическом исследовании, зачастую могут зависеть от человеческого фактора, связаны с риском из-за инвазивности и ограниченной доступности, а также требуют временных затрат. Развитие методов молекулярной генетики углубило понимание патогенеза глиобластом, однако в клинической практике по-прежнему есть потребность в поиске и отборе новых биомаркеров для ранней диагностики, в том числе для создания стандартизированных диагностических панелей, которые будут эффективно дифференцировать опухоли.

В качестве подобных биомаркеров перспективно изучение циркулирующих микроРНК – коротких некодирующих РНК, которые являются регуляторами генной экспрессии через пре- и посттранскрипционную активность [3] и сохраняют стабильность в биожидкостях, вследствие чего открывают перспективы для малоинвазивной диагностики.

Основная часть

Профили микроРНК являются высокоразмерными данными, которые могут содержать тысячи признаков при относительно небольших клинических выборках, поэтому важен правильный выбор алгоритмов отбора наиболее значимых микроРНК, позволяющих отличать злокачественные новообразования головного мозга между собой и здоровыми образцами.

Одним из решений уменьшения размерности пространства признаков является объединение статистического анализа дифференциальной экспрессии микроРНК между группами с машинным обучением для уменьшения биологического шума, выявления паттернов экспрессии биомаркеров и построения мультиклассовых моделей-классификаторов. Использование открытых данных секвенирования нового поколения обусловлено тем, что данные содержат более полный профиль микроРНК в отличие от кПЦР, где диапазон детектируемых микроРНК ограничен выбранными праймерами.

В данной работе был проведен анализ сырых данных секвенирования нового поколения 167 образцов плазмы крови пациентов разных диагностических групп

(здоровые образцы, метастаза глиобластомы (GBM) и метастаза головного мозга (MBM)) из открытой базы данных GEO NCBI в виде fastq файлов. Проведена предобработка сырых данных в виде удаления адаптерных последовательностей, фильтрации чтений по качеству и длине. Затем была получена матрица представленности микроРНК по всем образцам с использованием инструмента Mirge3. Дифференциальная экспрессия рассчитывалась с помощью пакета DESeq2. На основе данных экспрессии отдельных микроРНК был произведен отбор наиболее статистически значимых признаков, после чего были обучены модели на основе RF, XGBoost, LASSO, KNN и SVM.

Выводы

По результатам отбора по статистическим критериям $|\log_2 FC| > 2$ и $\text{padj} < 0.05$, было выделено 39 биомаркеров. На их основе были обучены несколько моделей для оценки их применимости к задаче классификации новообразований головного мозга. Наилучший общий результат по всем группам показали SVM и KNN, тогда как RF, XGBoost и LASSO оказались более чувствительны к менее представленным классам GBM и MBM, для которых показывали меньшую точность предсказания, в то время как точность классификации здоровых пациентов оставалась высокой. Это может быть связано с меньшим размером выборок GBM и MBM, а также со схожим профилем экспрессии микроРНК. Помимо этого, оценивалось пересечение значимых микроРНК для классификации в каждой из моделей. Дальнейший анализ значимости признаков показал, что 26 из 39 первоначально отобранных микроРНК имели вклад хотя бы в одной модели.

Высокая общая точность предсказаний и пересечение признаков подтверждает возможность применения 26 микроРНК в диагностических панелях как биомаркеров для разделения образцов с новообразованиями головного мозга и здоровых образцов.

Литература

1. Su F, Liu Y, Zong Y, Gao Z, Zhou G, Deng C, Liu Y, Zeng Y, Ma X, Wang Y, Wu Y, Xu F, Guan L, Liu B. Identification of circulating miRNA as early diagnostic molecular markers in malignant glioblastoma base on decision tree joint scoring algorithm. *J Cancer Res Clin Oncol*. 2023 Dec;149(20):17823-17836. doi: 10.1007/s00432-023-05448-w. Epub 2023 Nov 9. PMID: 37943358; PMCID: PMC11797510.
2. Tbahriti HF, Boukadoum A, Benbernou M, Belhocine M. Machine learning and deep learning in glioblastoma: a systematic review and meta-analysis of diagnosis, prognosis, and treatment. *Discov Oncol*. 2025 Aug 7;16(1):1492. doi: 10.1007/s12672-025-03303-7. PMID: 40773129; PMCID: PMC12331544.
3. Ordóñez-Rubiano EG, Rincón-Arias N, Espinosa S, Shelton WJ, Salazar AF, Cómbita A, Baldoncini M, Luzzi S, Payán-Gómez C, Gómez-Amarillo DF, Hakim F, Patiño-Gómez JG, Parra-Medina R. The potential of miRNA-based approaches in glioblastoma: An update in current advances and future perspectives. *Curr Res Pharmacol Drug Discov*. 2024 Jun 29;7:100193. doi: 10.1016/j.crphar.2024.100193. PMID: 39055532; PMCID: PMC11268206.