

Секция «Теория вероятностей и математическая статистика»

Учет редких мутаций при поиске генных комбинаций, влияющих на уровень окисленных липопротеидов низких плотностей

Шевцов Антон Евгеньевич

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова,

Механико-математический факультет, Кафедра теории вероятностей, Москва, Россия

E-mail: Shevan05@gmail.com

Определение нуклеотидной последовательности ДНК проводится с помощью секвенирования с целью выявления мутаций, приводящих к возникновению полиэтиологических заболеваний. Одна из проблем, возникающих в таких исследованиях, учет вклада редких мутаций в развитие заболевания. Решение этой проблемы осложняется низкой частотой встречаемости таких мутаций в генотипе человека и требует большого объема выборки, что не всегда предоставляется возможным при планировании исследования. По этой причине были развиты математические методы, позволяющие понять, является ли заболевание или его отсутствие следствием наличия группы некоторых мутаций. При этом данные методы, позволяют анализировать интегральный эффект группы генов на наличие/отсутствие заболевания. Наиболее распространенными среди них являются: CAST («Cohort Allelic Sum Test») [1], СМС («Combined Multivariate and Collapsing») [2], WSS («Weighted Sum Statistic») [3]. Первые два метода основаны на интегральном подходе, позволяющем объединять разными способами группы генов в одну, и дальнейшем анализе полученных суммарных данных многомерными статистическими тестами (например, тестом Хотеллинга T²). С помощью метода WSS на основе данных по всей выборке формируются веса для каждой мутации. Сумма таких весов по всем генам и пациентам представляет собой случайную величину, и затем проверяется гипотеза о том, что эта величина распределена нормально. Как показано в [3], мощность статистического теста WSS превышает мощность тестов, предложенных в [1,2]. Цель настоящей работы [U+2500] применить эти методы к данным, полученным в лаборатории молекулярной генетики ФГБУ «Государственный научно-исследовательский центр профилактической медицины» МЗ РФ. Данные представляют собой наборы однонуклеотидных полиморфизмов (SNP или мутаций) и уровни окисленных липопротеидов низких плотностей (oxLDL) в плазме крови каждого пациента из выборки пациентов (n=717), с низким риском развития сердечно-сосудистых осложнений. В работе уровень oxLDL сводился к бинарному параметру и считался повышенным при значениях, превышающих выборочную медиану oxLDL. Использование методов CAST и СМС осложнялось немногочисленностью рассматриваемой группы генов. При этом метод WSS оказался применим в данном исследовании, и с его помощью было выявлено наличие статистической зависимости между интегральным показателем, характеризующим рассматриваемую группу генов, и уровнем oxLDL. Анализ данных проводился с помощью программного обеспечения R для статистической обработки данных. Рассматриваемые подходы помогут заложить базу для создания нового метода по анализу редких мутаций.

Источники и литература

- 1) 1. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, et al. (2004) Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 305: 869–872.
- 2) 2. Li B, Leal SM (2008) Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* 83: 311–321.

- 3) 3. Madsen BE, Browning SR (2009) A Groupwise Association Test for Rare Mutations Using a Weighted Sum Statistic. PLoS Genet 5(2): e1000384. doi:10.1371/journal.pgen.1000384