

Новый инструмент определения дифференциальной экспрессии аллелей ГКГС

Научный руководитель – Шарова Елена Ивановна

Гончаров Евгений Алексеевич

Студент (бакалавр)

Московский физико-технический институт, Москва, Россия

E-mail: goncharov.ea@phystech.edu

Аллели главного комплекса гистосовместимости (ГКГС) играют важную роль в иммунном ответе и связаны с предрасположенностью к заболеваниям. Их вариации определяют, как иммунная система распознает патогены и собственные клетки. Особенно важно проводить анализ генов ГКГС из данных РНК-секвенирования. В отличие от ДНК-секвенирования, такого типа данные гораздо более информативны, но имеют особенности, влияющие на точность типирования ГКГС.

Однако, помимо качественного результата стоит иметь оценку уровня экспрессии аллелей внутри каждого гена. Существует ряд инструментов типирования аллелей ГКГС, при этом только HLA-HD работает на геномных, экзомных и транскриптомных данных с наибольшей точностью, и только arcasHLA заявлен как дающий оценку уровня экспрессии [1, 3]. Так как arcasHLA не универсален и сложен в получении результата, мы разработали новый протокол вычисления образец-специфичной экспрессии аллелей, использующий результаты типирования HLA-HD, а в перспективе любого подобного инструмента, работающего на транскриптомных данных.

Для оценки результативности подхода были взяты 3 набора данных РНК-сек с числом прочтений не менее 20 млн в каждом: Geuvadis (клетки с высокой экспрессией всех классов ГКГС, 179 образцов) и образцы ткани кишечника (77 от пациентов с болезнью Крона, 43 от пациентов без патологий). На них мы реализовали наш новый подход расчета экспрессии аллелей с помощью инструмента Salmon [4] и образец-специфичного набора транскриптов аллелей ГКГС.

Выявлено, что часть аллелей, обнаруженных HLA-HD, являются ложными (артефактными) аллелями с TPM ~ 0 , а экспрессия истинных аллелей находится в диапазоне 40-1500 TPM. Также было обнаружено, что артефакты типировались тканеспецифично (Рис. 1). Аллель DRB3*02:96 ранее был отмечен как артефактный в экзомных данных [2], в транскриптомных данных он не обнаружен в наборе Geuvadis и обнаруживается в тканях кишечника. При этом аллель, на основе которого обнаруживается артефакт, присутствует во всех анализируемых наборах данных. Также мы впервые обнаружили возможную артефактность типирования аллеля DMB*01:05 (Рис. 2). То есть наш подход обеспечивает эффективную оценку аллель-специфичной экспрессии, в том числе выявляя артефакты типирования ГКГС.

Данная работа выполнена в рамках государственного задания «Т-клетки», номер государственного учета НИОКТР 123032900030-7

Источники и литература

- 1) Claeys A. [и др.]. Benchmark of tools for in silico prediction of MHC class I and class II genotypes from NGS data // BMC Genomics. 2023. № 1 (24). С. 247.

- 2) Dashti M. [и др.]. Evaluation of HLA typing content of next-generation sequencing datasets from family trios and individuals of arab ethnicity // *Frontiers in Genetics*. 2024. (15). С. 1407285.
- 3) Orenbuch R. [и др.]. arcasHLA: high-resolution HLA typing from RNAseq // *Bioinformatics*. 2020. № 1 (36). С. 33–40.
- 4) Patro R. [и др.]. Salmon provides fast and bias-aware quantification of transcript expression // *Nature Methods*. 2017. № 4 (14). С. 417–419.

Иллюстрации

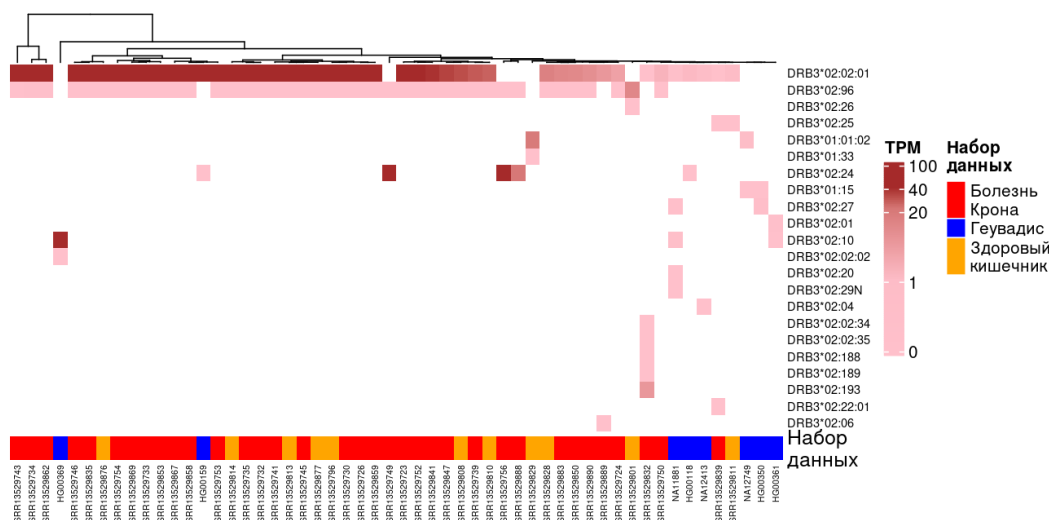


Рис. : Рис. 1. Тканеспецифичное типирование аллеля DRB3*02:96

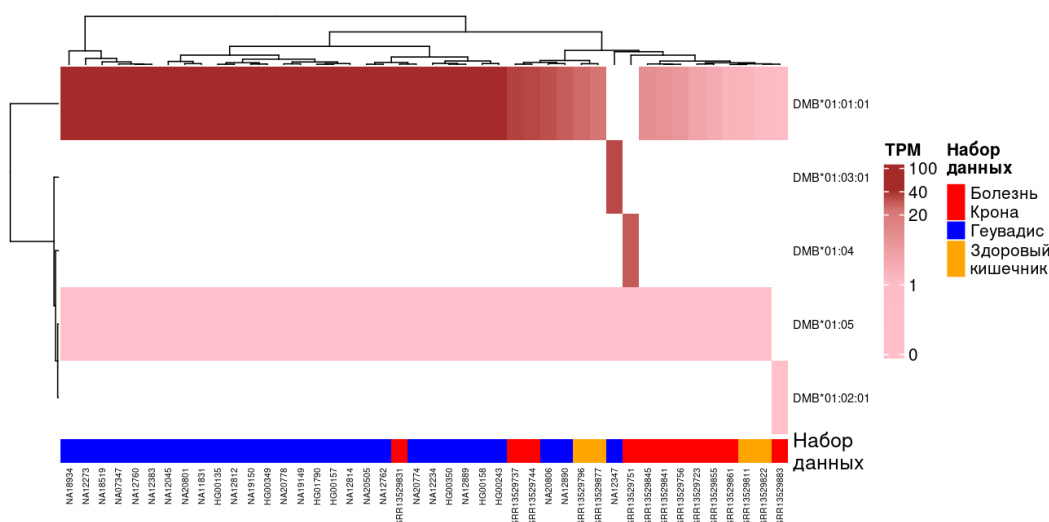


Рис. : Рис. 2. Артефактность типирования DMB*01:05