Секция «Биоинформатика»

Глубокая модель селекции Т-клеточных рецепторов

Научный руководитель – Горбачев Дмитрий Викторович

Офицеров Евгений Петрович

Аспирант

Тульский государственный университет, Тула, Россия

E-mail: eofitserov@gmail.com

Адаптивный иммунитет играет важнейшую роль в защите организма от различных патогенов. В настоящее время благодаря развитию технологий секвенирования появилась возможность извлекать данные о нуклеотидных последовательностях, кодирующих Т-клеточные рецепторы, и использовать их для количественного моделирования иммунной системы.

Ранее в работе [2] была предложена модель тимической селекции Т-клеточных рецепторов. В ней вероятность селекции предсказывается на основе длины аминокислотной последовательности, уникальной комбинации VJ генов и распределения аминокислот по позициям в последовательности. Параметры модели выводятся на основе данных о распределении Т-клеточных рецепторов, полученных с помощью многопоточного секвенирования, а также вероятностной модели процесса V(D)J рекомбинации, описанной в работе [1]. В предложенной модели делается предположение об условной независимости аминокислот в гипервариабельной части рецептора. Это является ее существенным недостатком.

В данной работе предлагается более точная контекстно-зависимая модель тимической селекции на основе глубоких нейронных сетей. Кроме того, разработан эффективный алгоритм обучения этой модели на основе экспериментальных данных.

В предлагаемой модели вероятность того, что рецептор пройдет селекцию, описывается с помощью двунаправленной рекуррентной нейронной сети. Отметим, что рекуррентные сети зарекомендовали себя как эффективный подход при описании последовательностей, позволяя учитывать контекст произвольного размера. Для вывода параметров такой модели, также как в работе [2], использовалось два набора рецепторов: репертуар, полученный с помощью секвенирования, и выборка из 3,000,000 случайных последовательностей, полученных с помощью вероятностной модели V(D)J рекомбинации (см. [1]). Обучение модели заключается в максимизации правдоподобия экспериментальной выборки. Для использования при обучении стандартных градиентных методов был разработан подход, в котором для оценки нормализующей константы (см. [2]) выполняется повторное семплирование по Гиббсу из множества сгенерированных последовательностей. Такой подход позволяет обучать различные архитектуры нейронных сетей и значительно лучше предсказывать давление селекции на различные рецепторы.

Источники и литература

1) 1. Anand Murugan et al. 2012. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. Proceedings of the National Academy of Sciences, 109, 40, 16161–16166. Jorge Nocedal, Stephen J. Wright. 2006.
2) 2. Yuval Elhanati et al. 2014. Quantifying selection in immune receptor repertoires. Proceedings of the National Academy of Sciences,111, 27, 9875–9880.