

**Генерация регуляторных последовательностей дрожжей с заданным влиянием на уровень экспрессии при помощи методов глубокого обучения**

**Носкова Елизавета Олеговна**

*Студент (специалист)*

Московский государственный университет имени М.В.Ломоносова, Факультет  
биоинженерии и биоинформатики, Москва, Россия

*E-mail: chemistrykekulen@gmail.com*

В большинстве генетических конструкций используют небольшой набор промоторов. Возможность создания синтетических промоторов с заданным уровнем экспрессии генов может облегчить множество задач генной инженерии и биотехнологии путем тонкой настройки уровней экспрессии генов. В частности, это может использоваться для определения влияния уровня экспрессии генов интереса на фенотип и для персонализации генной терапии.

Чтобы решить такого рода задачу для сложных многоклеточных организмов, сначала стоит обратиться к более простым модельным объектам, таким как, пекарские дрожжи *Saccharomyces cerevisiae*.

В нашей работе мы создали нейронную сеть, основанную на архитектуре сверточной сети EfficientNetV2, для генерации промоторов *S. cerevisiae* с заданными свойствами. В качестве обучающей выборки мы использовали данные полученные в результате масштабного эксперимента по измерению активности произвольных промоторных последователей дрожжей *S. cerevisiae* [n1].

Мы решили взять за основу подход, используемый в современных генеративных нейронных сетях, основанных на диффузии. На этапе обучения мы вносим случайные мутации в последовательность и просим нейросеть исправить эти мутации, используя мутированную последовательность, экспрессию исходной последовательности и число внесенных мутаций.

Генерация последовательностей происходит путем итеративного снижения уровня шума. На вход обученной нейронной сети подается произвольная последовательность и целевой уровень экспрессии. Далее выбирается количество мутаций, которые должна исправить модель. Сгенерированная последовательность снова подается модели пока из случайного шума не будет получена последовательность с экспрессией близкой к целевой. Для контроля качества сгенерированных последовательностей используется другая нейронная сеть с аналогичной архитектурой, обученная определять экспрессию промоторных последовательностей. Как метрика качества используется средний квадрат отклонения между предсказанной экспрессией сгенерированной последовательности (MSE) и целевой экспрессией, а также корреляция Спирмена и Пирсона. Таким образом, у нас получилось сгенерировать последовательности с корреляцией предсказанной и желаемой экспрессии Спирмена 0.883 и Пирсона 0.869 и MSE 3. Это значительно больше, чем корреляция полученная в аналогичной работе с использованием генеративно-состязательной сети, где корреляция Пирсона составила 0.514 [n2].

### **Источники и литература**

- 1) Carl G. de Boer, et al. "Deciphering Eukaryotic Gene-Regulatory Logic with 100 Million Random Promoters." *Nature Biotechnology* 38, no. 1 (January 1, 2020): 56–65.
- 2) Zrimec Jan, et al. "Controlling Gene Expression with Deep Generative Design of Regulatory DNA." *Nature Communications* 13, no. 1 (August 30, 2022): 5099.