*Тестовое задание в Par Seq Антонова Екатерина*

**Задача №1**

Воспользуемся аннотацией генома человека - файл **gencode.v19.annotation.gff3**

Находим перекрывающееся регионы в аннотации генома и панели – сохраняем результат в новый файл (команда терминала Linux):

**bedtools intersect -a gencode.v19.annotation.gff3 -b** **IAD143293\_241\_Designed.bed > annotation\_panel.bed.**

Далее запускаем программу, работающую с полученным файлом, для вывода генов и номеров их экзонов для которых проводится обогащение **find\_exons&genes.py**. Формат результата работы программы в формате .txt :

*Ген SLC16A1, экзон(ы) [1, 2, 3, 4, 5]*

*Ген RP11-31F15.1, экзон(ы) [1]*

*Ген KLF11, экзон(ы) [2, 3]*

*. . .*

Далее запускаем вторую программу **ID\_maker.py** которая добавляет в изначальный файл два столбца: название гена, название экзона. Полученный результат записывается в IAD143293\_241\_Designed\_with\_IDs.bed

*chr1 113460270 113460593 AMPL7163702396 Pool=2;SUBMITTED\_REGION=AMPL7163702396 SLC16A1 2*

*. . .*

Анализ генов проводим на сайте **https://www.ncbi.nlm.nih.gov/gtr/tests**

Данная панель может быть использована для диагностики группы генетических эндокринных заболеваний, связанных с нарушением метаболизма глюкозы. Это и врожденный гиперинсулизм, перманентный неонатальный сахарный диабет, сахарный диабет молодого типа с началом в зрелом возрасте, диабет и ожирение и т.д.

**Задача №2**

С помощью **https://meme-suite.org/meme/tools/bed2fasta** преобразуем изначальный файл в формат .fasta (output\_meme.fa). Загружаем его в **BLAST**,выбираем геном человека RefSeqGene и скачиваем результаты в формате .csv (ZHWCBC5W013-Alignment-HitTable.csv) и запускаем программу **find\_mismatches.py**

На выходе получаем строки, в которых указаны идентификаторы последовательностей из панели и генов, на которые могут неправильно выровняться последовательности или которые могут ошибочно амплифицироваться в ходе анализа.

Ответ:

chr2:169765146-169765432(+) NG\_011682.1 NG\_027759.1 NG\_023267.2 NG\_011570.3 NG\_033956.2 NG\_052859.1 NG\_047135.2

chr2:169765412-169765633(+) NG\_011682.1 NG\_008475.1 NG\_029628.2 NG\_028282.1 NG\_011463.1 NG\_008448.1

chr6:29640272-29640598(+) NG\_031873.1 NG\_013045.1

chr10:88820672-88820809(+) NG\_016456.1 NG\_013010.1

chr11:2180826-2181029(+) NG\_050578.1 NG\_007114.1

chr11:17414432-17414602(+) NG\_012446.1 NG\_008867.1

chr12:121440298-121440554(+) NG\_011731.2 NG\_051308.1 NG\_009165.3 NG\_032009.1 NG\_013218.1 NG\_012576.1

chrX:49107143-49107461(+) NG\_021311.2 NG\_007392.1

. . .