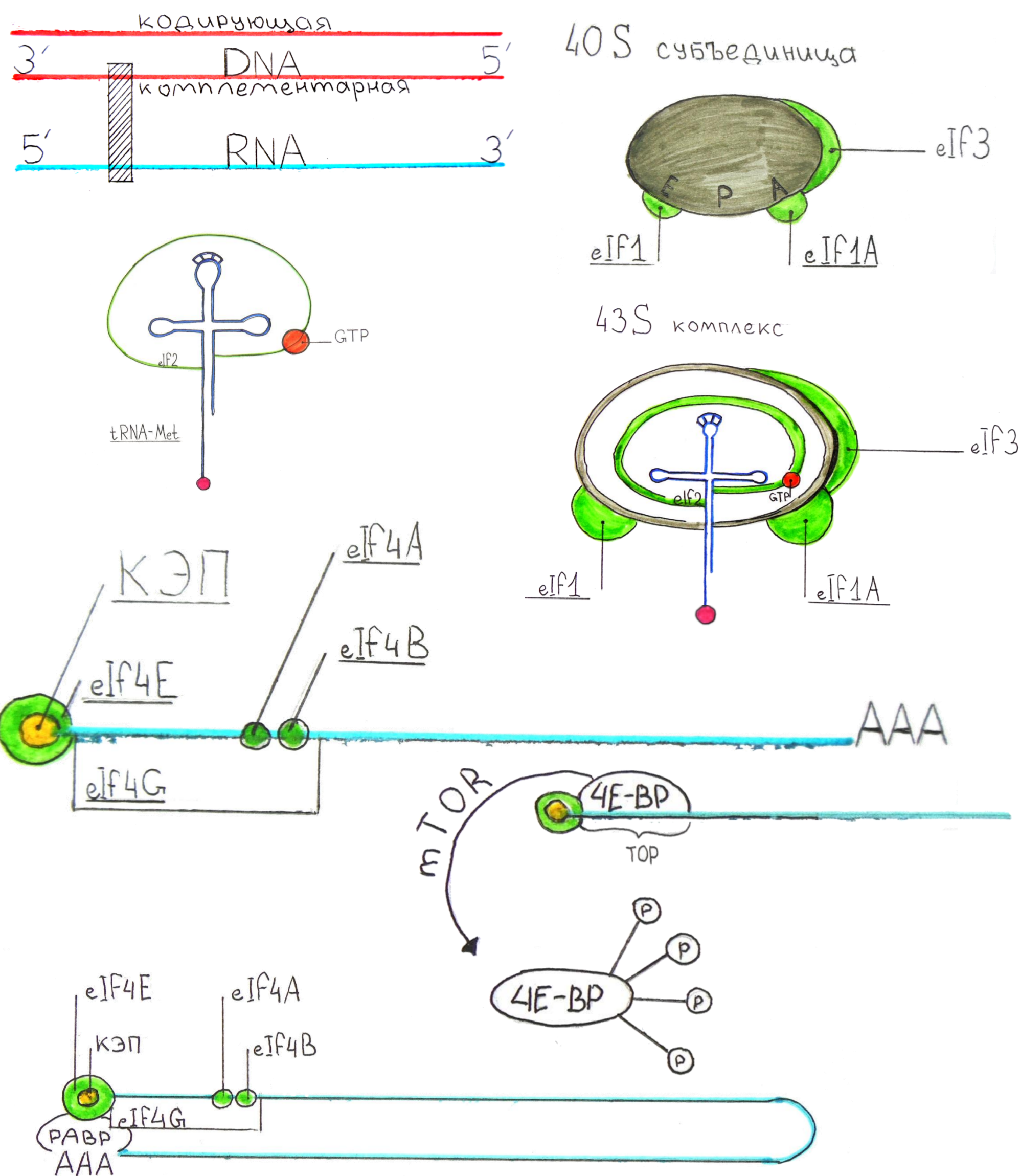


Компьютерный анализ регуляторных последовательностей 5' UTR mRNA-мишеней mTOR-сигнального каскада



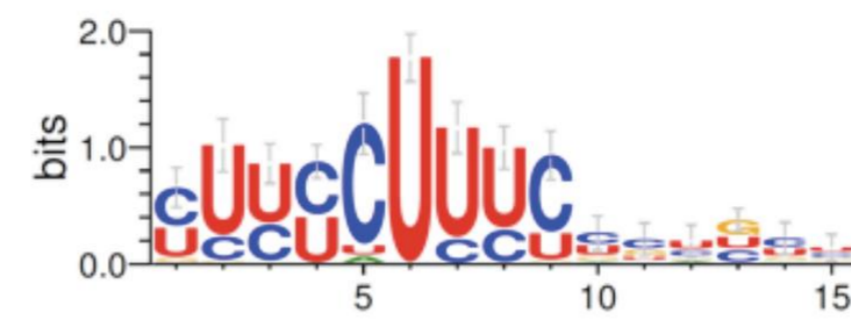
Кирилл Бабеев, Софья Буюнова, Мария Сысоева
Руководитель проекта: Иван Кулаковский Научный консультант: Ирина Елисеева

Лаборатория регуляторной геномики

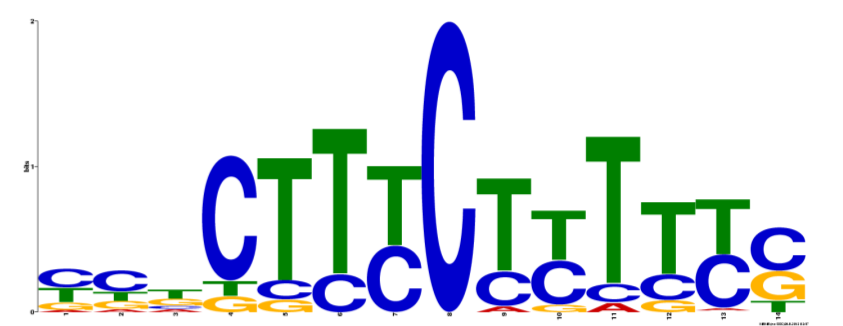


[Hsieh и др., Nature, 2012, PubMed ID: 22367541]

Pyrimidine Rich Translational Element (PRTE)

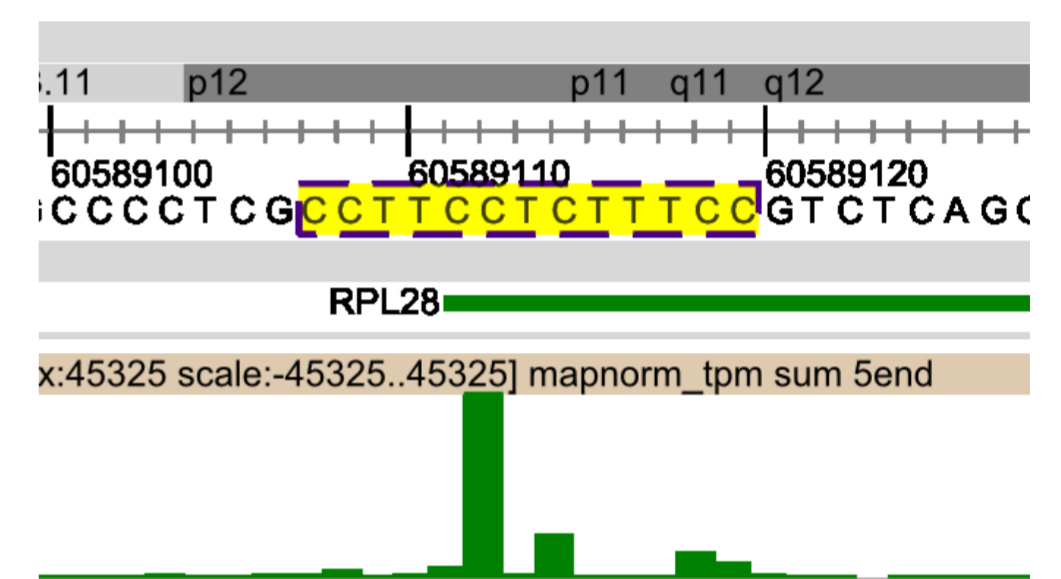


Сборка генома hg18,
Мотив найден в 97 из 144 5'UTR



Сборка генома hg19,
Мотив найден в 110 из 144 5'UTR

В ~20 случаях мотив найден в начале последовательности
- возможно ли, что это просто длинный TOP-мотив?



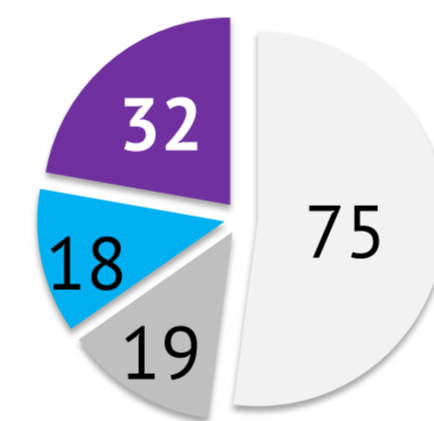
Из статьи [Hsieh и др., Nature, 2012, PubMed ID: 22367541] мы знаем о предположительной структуре 5' UTR mTOR-зависимых mRNA:

TOP (terminal oligopyrimidine tract) - короткая СТ-богатая лидерную последовательность, состоящая из 5-ти и более нуклеотидов в начале 5' UTR

PRTE (pyrimidine-rich translational element) - СТ-богатая последовательность, состоящая из 10 и более нуклеотидов в середине и конце последовательности 5' UTR

Кроме того, авторами выдвигалось предположение, что PRTE-не имеет позиционных предпочтений по локализации в 5' UTR.

TORTE – Terminal Oligopyrimidine Regulatory Translational Element



- Найден TORTE > 10 нуклеотидов непосредственно над мажорным стартом (RPL28)
- Найден небольшой TORTE (6-10 нуклеотидов) над мажорным стартом (EEF2)
- Найден TORTE над альтернативным стартом (RPS25)
- Не найден выраженный TORTE (NCLN)

Поиск TOP-мотива с помощью регулярных выражений

модель	лидер 20 нукл., тестовая выборка	лидер 20 нукл., все геномные mRNA
минимум 5 букв, ТОЛЬКО СТ. [CT]+	78 из 144	20471 из 59911
минимум 6 букв, 1 допустимая вставка A или G [CT]+[AG]?[CT]+	94 из 144	27632 из 59911

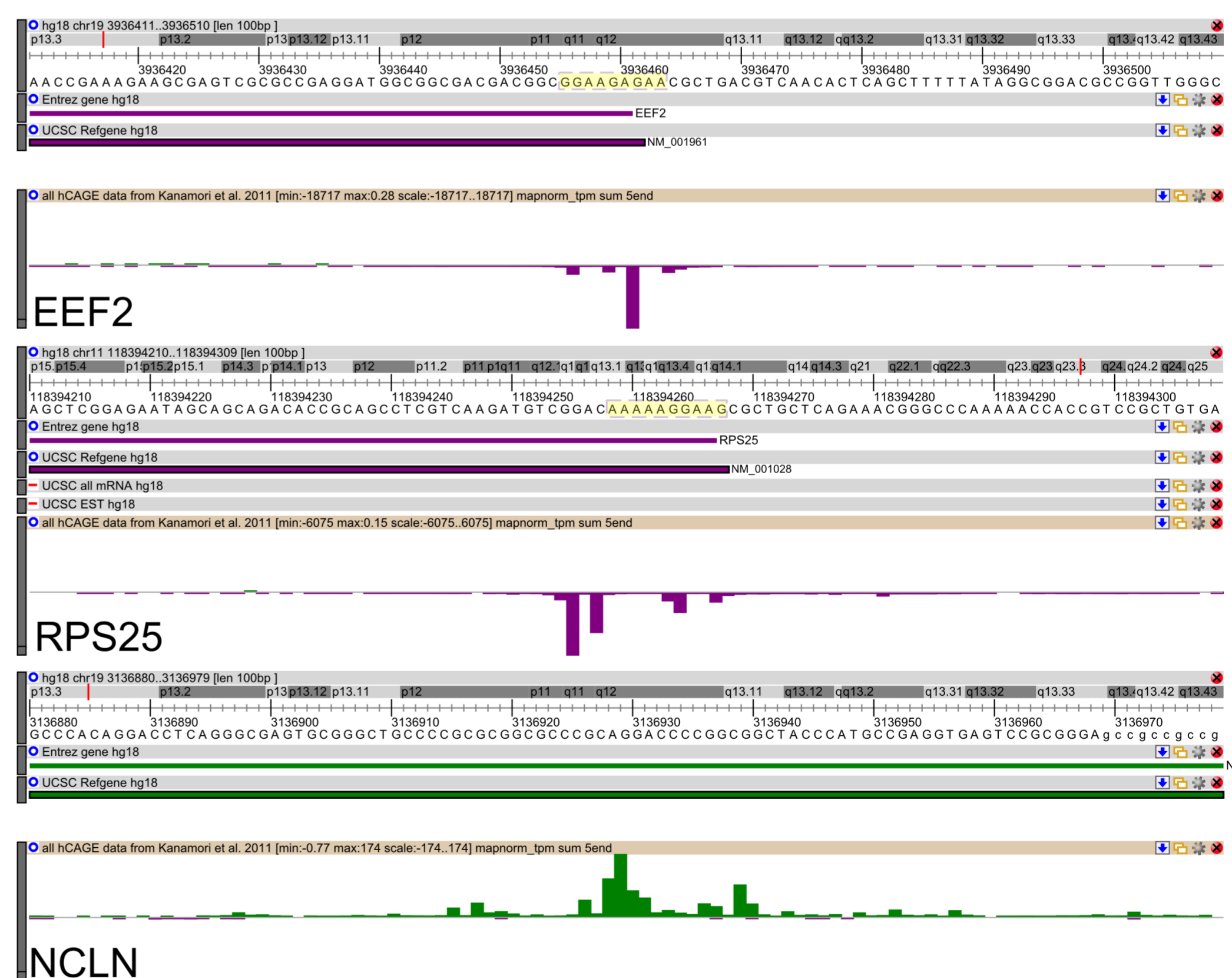
Частота TOP-содержащих последовательностей
94/144 = **0.65** в тестовой выборке;
27632/59911 = **0.46** в пологономном наборе mRNA.

Поиск PRTE-мотива с помощью регулярного выражения

Взяты «хвосты» последовательностей (без лидера 20)	Число PRTE-содерж. последов-й	Слишком короткие UTR (<20+12 нукл.)
Тестовая выборка	90 из 144	1 из 144
Все геномные mRNA	33244 из 59911	10800 из 59911

Регулярное выражение /[CT]+[AG]?[CT]+[AG]?[CT]+/
Дополнительно контролировали длину (не менее 12 букв).

Частоты PRTE-содержащих последовательностей:
90/(144-1) = **0.63** в тестовой выборке;
33244/(59911-10800) = **0.68** в геномных mRNA;
PRTE в тестовой выборке встречается реже, чем в геноме?



Оценка точности картирования стартов транскрипции по выборке mTOR-зависимых mRNA

Старт в базе данных картирован с точностью:	Число последовательностей
- до 10 нуклеотидов	63
- от 11 до 100 нуклеотидов	70
- 101 и более нуклеотидов	11