

Анализ корреляции между биохимическими свойствами ампликонов и эффективностью амплификации в мультиплексной ПЦР

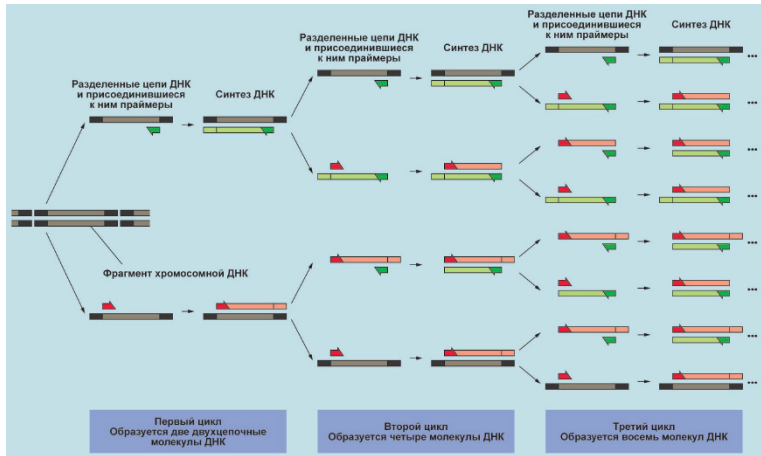
Студент: Татьяна Кривошеева
Руководители проекта: Антон Брагин¹, Герман Демидов^{1,2}

¹Parseq Lab

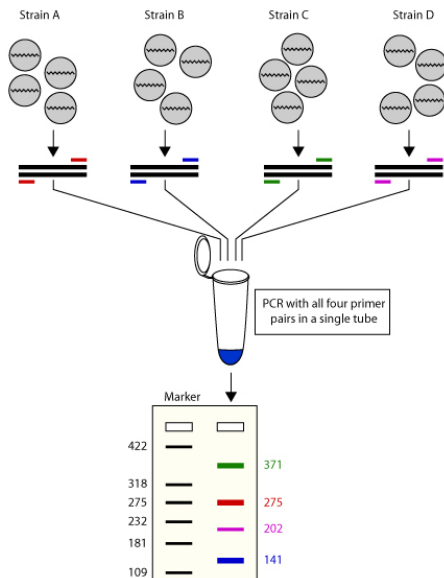
²СПбАУ

Презентации проектов, осенний семестр 2014

ПЦР



Мультиплексная ПЦР



Задачи

- Определить набор свойств, от которых может зависеть эффективность ампликонов
- Понять, от каких свойств эффективность зависит в большей степени
- Кластеризовать ампликоны по поведению в серии реакций и определить, какие свойства влияют на попадание ампликона в кластер
- Предложить подходы к оценке качества работы мультиплексных панелей и подбору праймеров

Задачи

- Определить набор свойств, от которых может зависеть эффективность ампликонов
- Понять, от каких свойств эффективность зависит в большей степени
- Кластеризовать ампликоны по поведению в серии реакций и определить, какие свойства влияют на попадание ампликона в кластер
- Предложить подходы к оценке качества работы мультиплексных панелей и подбору праймеров

Данные

- два набора данных
- каждый включает в себя информацию о количестве ампликонов таргетной панели в 100 – 150 образцов
- данные получены с использованием Ion Torrent PGM

Используемые программные средства

- R, RStudio
- Используемые библиотеки: FSelector, cvTools, randomForest, vegan

Признаки

- длина (праймера, ампликона)
- dangling ends (примыкающие к последовательности праймера нуклеотиды)
- GC-content (праймера, ампликона)
- последовательность (хэш от последовательности)
- ΔG_0 гибридизации праймера
- эффективная ΔG_0 гибридизации праймера (с учетом существования в растворе альтернативных конформаций)
- температура отжига
- нуклеотид 3'-конца
- эффективная концентрация дуплекса праймер-матрица
- энтропия последовательности ампликона

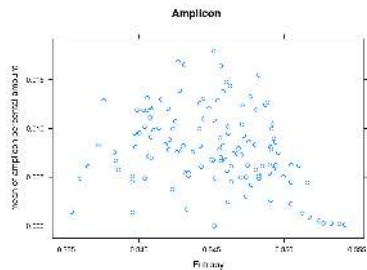
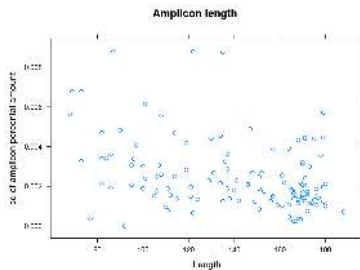
Подбор признаков. Доля

	χ^2	H	RF	BFS
Amplicon Len	●	●	●	●
Fwd 3'	●	●		
Rev 3'	●	●		
Fwd Primer Len			●	
Fwd Primer GC-content	-	-		●
Fwd Primer Entropy	-	-	●	
Rev Primer Len				
Rev Primer GC-content	-	-		
Rev Primer Entropy	-	-		
Amplicon GC-content	-	-	●	
Amplicon Entropy	-	-	●	●
Amplicon MaxPolyX				

Подбор признаков. Стандартное отклонение

	χ^2	H	RF	BFS
Amplicon Len	●	●	●	●
Fwd 3'	●	●		
Rev 3'	●	●		
Fwd Primer Len				
Fwd Primer GC-content	-	-		●
Fwd Primer Entropy	-	-	●	
Rev Primer Len				
Rev Primer GC-content	-	-		
Rev Primer Entropy	-	-		
Amplicon GC-content	-	-	●	
Amplicon Entropy	-	-	●	●
Amplicon MaxPolyX				●

Длина ампликона и энтропия его последовательности



Итоги

- Длина ампликона и энтропия последовательности ампликона влияют на среднее и стандартное отклонение
- Остальные признаки не влияют существенно на долю ампликона

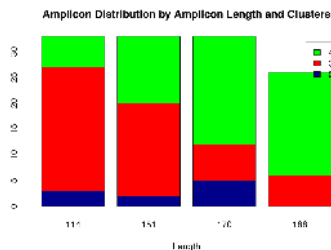
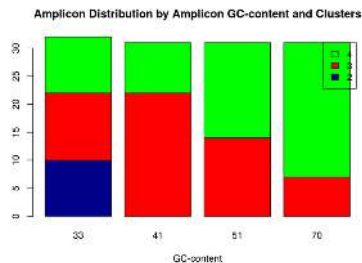
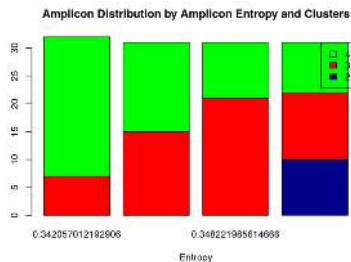
Задачи

- Определить набор свойств, от которых может зависеть эффективность ампликонов
- Понять, от каких свойств эффективность зависит в большей степени
- Кластеризовать ампликоны по поведению в серии реакций и определить, какие свойства влияют на попадание ампликона в кластер
- Предложить подходы к оценке качества работы мультиплексных панелей и подбору праймеров

Методы кластеризации

- KMeans
- Иерархическая кластеризация
- Иерархическая кластеризация корреляций

Распределение кластеров по значениям признаков



Итоги

Признаками, влияющими на поведение ампликона в ряде экспериментов, являются

- Энтропия последовательности ампликона
- GC-состав ампликона
- Длина ампликона

Спасибо! =)