

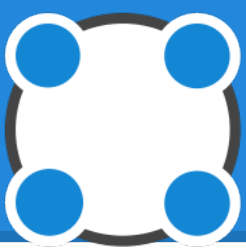


Моделирование ПЦР с высокой степенью мультиплексности на основе анализа данных таргетного NGS

Бушманова Вера, СПбГУ



Руководители: Антон Брагин, к.б.н.,
Герман Демидов, Parseq Lab, СПбАУ



Цель проекта

Построить хорошую модель мультиплексной ПЦР

Актуальность: диагностика гетерозиготных делеций в генах

CFTR, PAH, GALT, мутации в которых вызывают

муковисцидоз, фенилкетонурию и галактоземию,

соответственно.



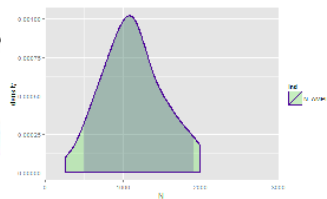
Задачи

- Изучение некоторых существующих моделей ПЦР и их применимости, улучшение этих моделей
- Построение модели, обладающей зависимостями, схожими с данными эксперимента (Ion-Torrent-e) и поиск характерных параметров
- Тестирование выбранной модели на реальных данных
- Получение плотности распределения начального количества ампликонов N_0 по параметрам ПЦР

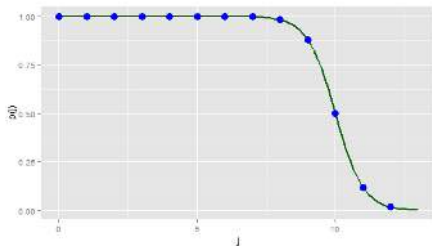
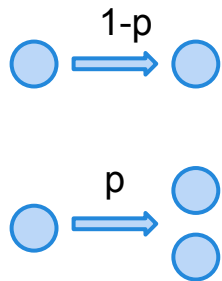


Модели

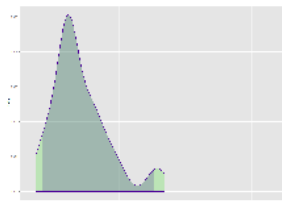
$$N_{j k} = (N_{j k}^0 + \delta_{j k}) \cdot c_0 \cdot (eff_j(\Omega) + \varepsilon_{1 j k}) \cdot (eff_j(\Omega) + \varepsilon_{2 j k}) \cdot \dots \cdot (eff_j(\Omega) + \varepsilon_{c_1 j k})$$



СТАТИСТИЧЕСКАЯ
МОДЕЛЬ



$$p = \frac{1}{1 + e^{(j-a) \cdot b}}$$



branching-модель

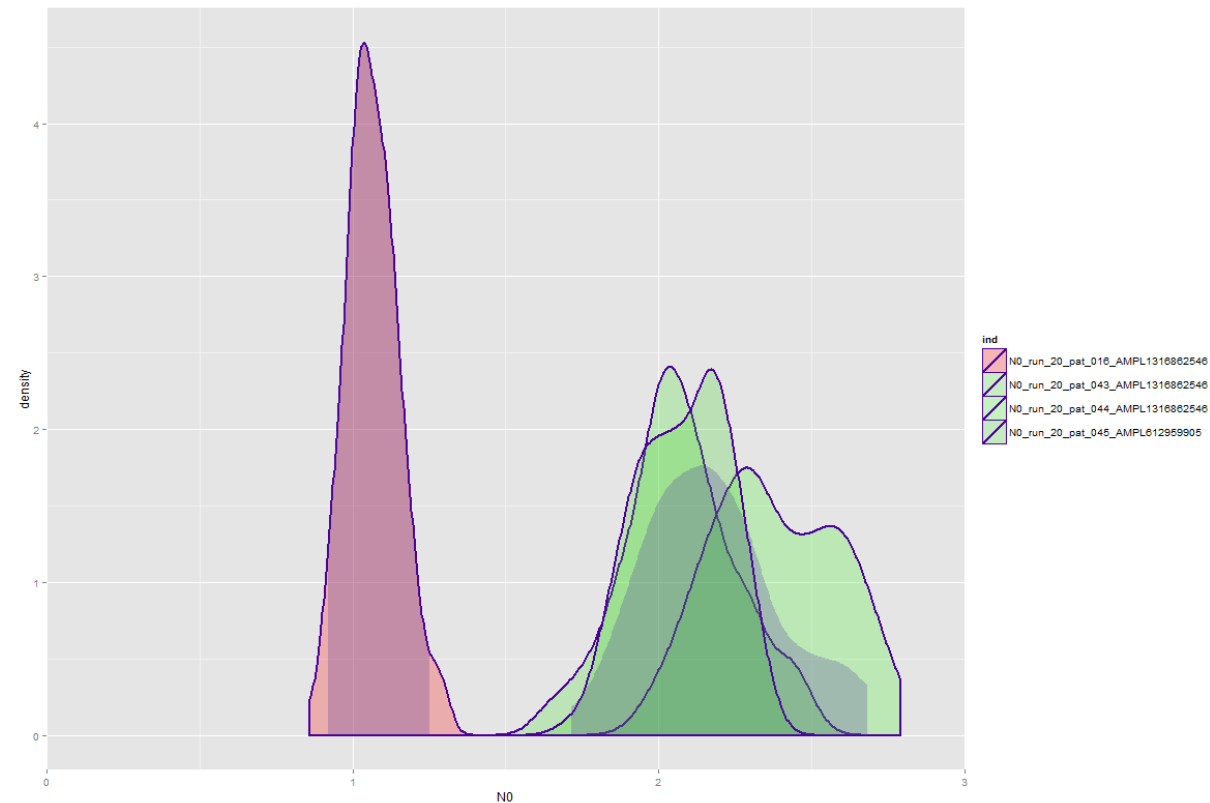
$$x_n = (1 + p)x_{n-1} + \tilde{x}_n$$

$$E\{x_n\} = E_{|x_{n-1}}\{E_{x_{n-1}}\{(1 + p)x_{n-1} + \tilde{x}_n\}\} = (1 + p)^n x_0$$

$$\sigma_n^2 = \frac{1-p}{1+p} [(1 + p)^{2n} - (1 + p)^n] x_0$$



Результаты

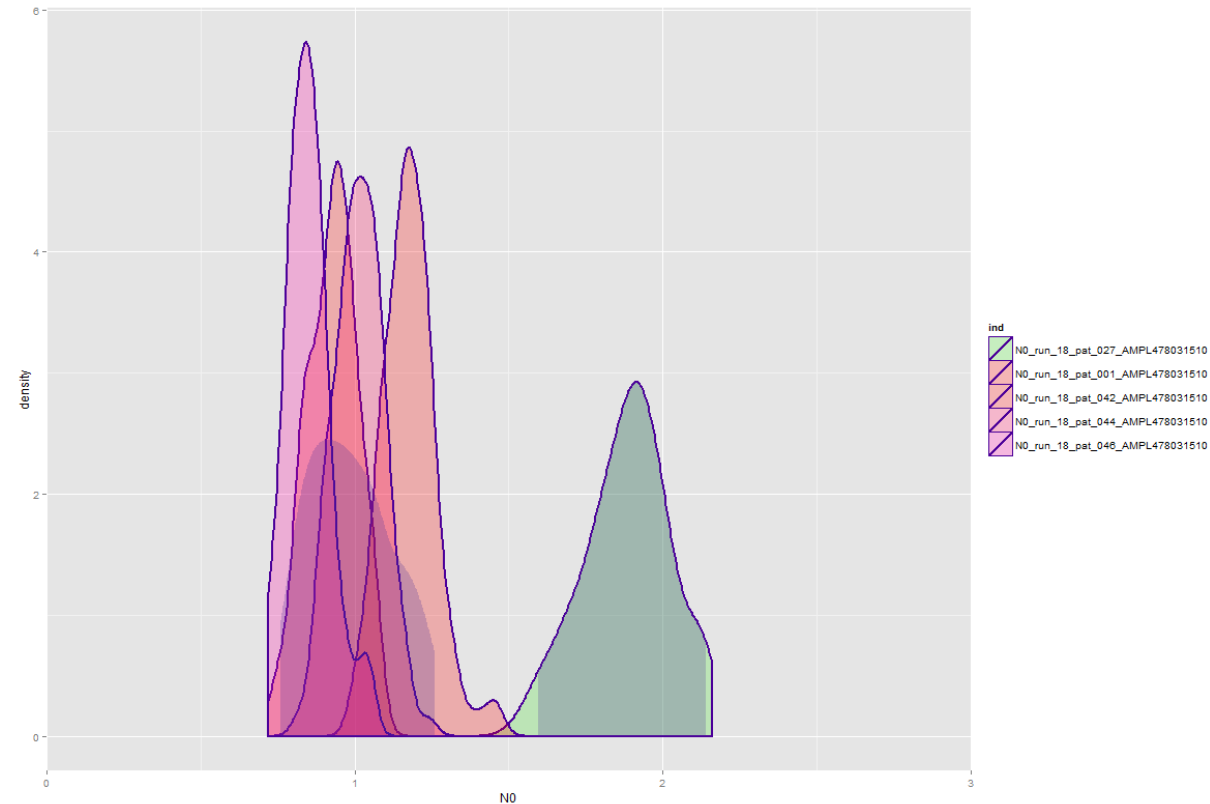


Плотность распределения начального количества ампликона (один ампликон, разные пациенты) Красной области соответствует больной пациент, зеленым - здоровые, обозначены квантили 0.025 и 0.975 (для здоровых вместе)

Полагаем, что процесс начинается с 2 ампликонов в случае без делеции и с 1 при делеции



Результаты



Несколько больных пациентов и один здоровый. Для больных для наглядности построен общий доверительный интервал.

Функции плотностей распределения хорошо разделяются на два типа: те, которые группируются около $N_0 = 1$ и $N_0 = 2$. Этому соответствуют делеция и нет.



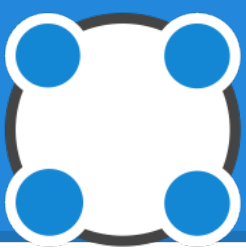
Обсуждение результатов

- Созданы статистическая и branching-модели для симуляции ПЦР
- Построены зависимости для реальных данных и для симулированных
- Эмпирически показано их соответствие
- Показана необходимость добавления гетероскедастичности в исходной приближенной формуле для статистической модели.



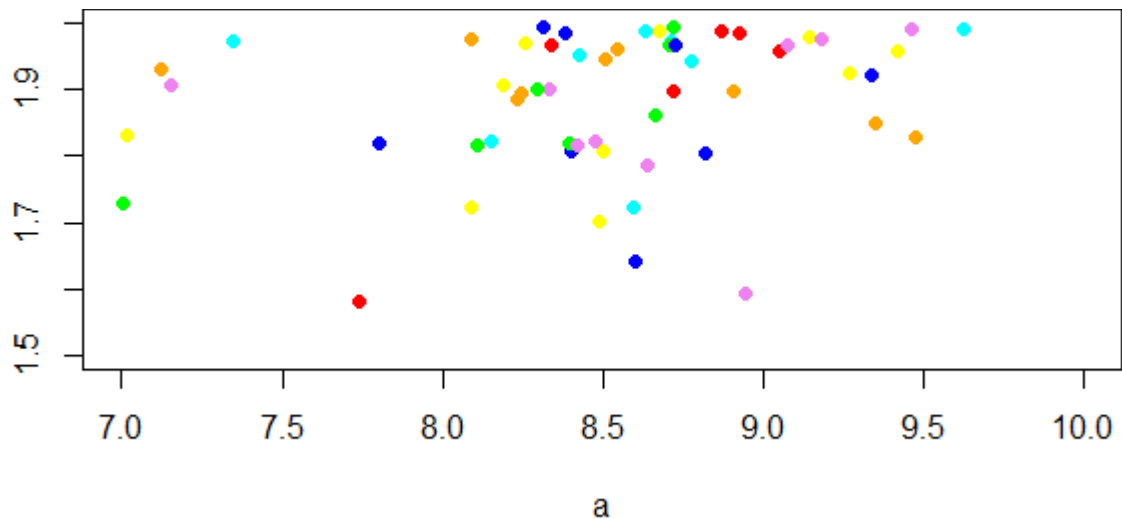
Обсуждение результатов

- Показано, что branching-процесс идёт не с постоянной вероятностью.
- Приближенная формула для вероятности в branching-процессе
- С помощью ML получены параметры ПЦР для реальных данных
- На основе вычисленных параметров ПЦР получены оценки для N_0



Сложности и планы

- Изменение процедуры нормализации
- Точность и ускорение сэмпирования по a и b
- Сэмплинг по a , b , N_0
- Разработка тула для детекции делекций b
- Сравнение с существующим решением





СПАСИБО ЗА ВНИМАНИЕ!

а также Герману, Антону, моей сестре Алёне и Диме!

<https://github.com/veranicebad>

