

- Тема работы.

Наследование способности к репрограммированию в клеточных делениях

- Состав коллектива: ФИО без сокращений, место работы/учёбы, учёные степени и звания.

Баттулин Нариман Рашитович, ИЦиГ СО РАН, НГУ, кбн.

Юнусова Анастасия Маратовна, ИЦиГ СО РАН

Фишман Вениамин Семенович, ИЦиГ СО РАН, НГУ, кбн

- Научное содержание работы:
 1. Постановка задачи.

В 2006 году биология стволовых клеток получила мощный импульс к развитию после того как японская группа под руководством Ш. Яманаки революционным образом изменила положение вещей в сфере технологий получения стволовых клеток. Эксперименты, проведенные авторами, показали, что фибробласты можно превратить в клетки, подобные эмбриональным стволовым при помощи введения в них четырех транскрипционных факторов (Oct4, Sox2, c-myc и Klf4). Полученные таким образом клетки авторы назвали индуцированными плюрипотентными стволовыми клетками (ИПС) клетками. Процесс перехода дифференцированной клетки в плюрипотентное состояние называется репрограммированием (Takahashi, Yamanaka, 2006). В настоящее время ИПС клетки считаются наиболее перспективным источником клеток для регенеративной медицины (обзор Robinton, Daley, 2012).

За последние несколько лет было опубликовано несколько тысяч исследований посвященных ИПС клеткам. Однако, несмотря на настоящий шквал работ, многие фундаментальные основы процесса репрограммирования остаются неизвестными. Среди наиболее важных проблем можно отметить низкую эффективность процесса репрограммирования. Из начальной популяции дифференцированных клеток лишь 0,01- 1% клеток дают начало полностью репрограммированным клонам ИПС клеток. Поэтому значительные усилия исследователей направлены на поиск способов увеличения эффективности получения ИПС клеток. На сегодняшний день предложен целый ряд способов позволяющих поднять эффективность репрограммирования, однако достичь революционного улучшения этого показателя пока не удалось. Возможно потому, что поиск факторов улучшающих репрограммирование вводится практически «в слепую»: до сих пор не известно, что отличает ту малую часть клеточной популяции, которая успешно превращается в ИПС клетки, от остальных клеток. Выявление закономерностей в распределении способности репрограммироваться в популяции эмбриональных фибробластов не только позволит сделать поиск факторов, увеличивающих эффективность репрограммирования, более эффективным, но также прольет свет на многие фундаментальные механизмы функционирования генома.

2. Современное состояние проблемы.

Для объяснения механизмов репрограммирования было предложено две альтернативные гипотезы: элитарная и стохастическая. Элитарная модель предполагает, что репрограммированию подвергаются только особая субпопуляция клеток, которая возможно представляет собой недифференцированные стволовые клетки, всегда присутствующие в первичных культурах. Стохастическая модель предполагает, что репрограммироваться может любая клетка, однако для успешного репрограммирования необходимо особое

физиологическое состояние – удачное сочетание многих случайно распределённых факторов. Получить доказательства в пользу стохастической модели удалось группе под руководством Рудольфа Яниша в элегантном эксперименте (Hanna et al., 2009). Для доказательства использовали линию «репрограммируемых» мышей (трансгенных мышей несущих в геноме трансгенную касету с факторами Яманакки под контролем Dox-индуцибельного промотора, что позволило авторам контролировать экспрессию факторов). Из мыши выделили популяцию незрелых В-лимфоцитов, рассадили по одной клетке в индивидуальные лунки и запустили экспрессию репрограммирующих факторов добавлением Dox. На протяжении следующих недель клетки делились и дали начало клеточному клону. Авторы следили за появлением репрограммированных клеток в каждом из клеточных клонов. И показали, что через 18 недель в 93% клеточных клонов появлялись ИПС клетки, а значит вся, или почти вся, исходная популяция, в принципе, может быть репрограммирована. Стоит особо подчеркнуть, что по стандартному протоколу колонии ИПС клеток формируются через 2-3 недели после старта экспрессии репрограммирующих факторов. В описываемом эксперименте формирование первых колоний ИПС клеток наблюдали также через 2 недели, тогда как последние появлялись через 18(!) недель. Это означает, что исходная В-клетка дала к этому времени клон, состоящий из многих миллионов клеток, из которых лишь единичные клетки дали начало ИПС клеткам. А исходную В-клетку от ИПС клеток разделял путь в 160 делений! Таким образом, было показано, что способность к репрограммированию может приобретаться в ходе клеточных делений.

С другой стороны при репрограммировании фибробластов человека было показано, что SSEA-3 субпопуляция фибробластов репрограммируется с большей эффективностью (Burn et al., 2009). Еще более убедительные доказательства справедливости элитарной модели, по крайней мере, в случае репрограммирования фибробластов человека были получены Вакао с соавторами. Авторам удалось показать, что в культуре фибробластов человека присутствуют мультипотентные стволовые клетки, названные Muse-клетки (Multilineage-differentiating stressenduring – мультипотентные стрессустойчивые клетки). При репрограммировании фибробластов человека ИПС клетки получались только из Muse-клеток, тогда как все остальные клетки были не способны к репрограммированию (Wakao et al., 2011). Таким образом ни одна из предложенных моделей не позволяет объяснить все имеющиеся экспериментальные данные.

В данной работе, мы собираемся ответить на актуальный в свете этого феномена вопрос: возможно ли закрепление свойства репрограммируемости и его наследование от материнской клетки дочерним в ходе деления.

3. Подробное описание работы, включая используемые алгоритмы.

В работе были использованы эмбриональные фибробласты мыши (ЭФМ) линии OG2 (ген Gfr под контролем промотора Oct4), любезно предоставленные д.б.н., чл.-корр. РАН, проф. А.Н. Томилиным (Санкт-Петербург). Согласно схеме эксперимента, эмбриональные фибробласты мыши были ко-трансдуцированы двумя лентивирусными конструкциями, одна из которых содержит репрограммирующие факторы Яманакки (Oct4, Sox2, Klf4 и c-Myc) под контролем доксициклин-индуцибельного промотора, а другая – индивидуальный 30-нуклеотидный ДНК-баркод и белок-трансактиватор. Именно этот трансактиватор при взаимодействии с доксициклином запускает экспрессию репрограммирующих факторов. Это позволяет нам контролировать экспрессию OSKM факторов добавлением доксициклина в культуральную среду. Клетки культивировались около двух суток до удвоения клеточной популяции, а затем были рассажены на 4 чашки, для того чтобы потомки одной клетки могли попасть в разные чашки. Затем, добавляя в среду доксициклин, мы запускали экспрессию репрограммирующих факторов. Таким образом, все баркодированные клетки вступали в процесс репрограммирования одновременно. После 7 дней индукции мы отбирали GFP-позитивные клетки, (GFP-индикатор репрограммирования), а их ДНК-баркоды секвенировали. Пересечение ДНК-баркодов с разных чашек свидетельствует о том, что потомки одной клетки попали на разные чашки и успешно репрограммировались. При этом, чем больше вероятность дочерних клеток получить способность к репрограммированию по наследству, тем большее количество пересечений ДНК-баркодов между чашками мы будем наблюдать. Массовое параллельное секвенирование ДНК-баркодов проводилось на платформе Ion Torrent в секторе геномных исследований ИЦиГ СО РАН в соответствии со стандартными протоколами производителя. Биоинформационный анализ данных

секвенирования проводился с помощью алгоритмов на языке Python. С помощью программы cutadapt 1.3 были удалены последовательности, фланкирующие ДНК-баркод, а также риды, длина которых не соответствовала заданной. Риды с низким качеством прочтения (где 20% позиций имеют значение Phred ниже 20), были удалены. Оставшиеся риды были переведены в FASTA формат и кластеризованы по гомологии с помощью программы DNAClust с пороговым значением гомологии 80%. Кластеры, представленные одним ридом, были исключены из анализа.

Мы разработали компьютерную модель, симулирующую все стадии эксперимента и позволяющую учесть ряд экспериментальных параметров. Первый шаг симулирует рассадку клеток и трансдукцию их вирусами. Каждая клетка в модели представлена как независимый объект. Эффективное число клеток N_0 считается как: $N_0 = N_{seeded} * (1 - e^{-m})$ где N_{seeded} это число клеток, посаженных на чашку в начале эксперимента, а m – MOI, или среднее количество вирусных интеграций на клетку. Этот параметр рассчитывается только для баркодированного вируса (LeGO-M2rtTA) из дополнительных экспериментов (см. результаты и методы ниже «Репрограммирование эмбриональных фибробластов»). Каждой клетке из N_0 присваивается ДНК-баркод. ДНК-баркод представляет собой случайное целое число, выбранное из количества вариантов, равного разнообразию библиотеки баркодов. В нашем случае это разнообразие составляет от 50 000 до 8 000 000 уникальных вариантов (см. результаты). Следующий шаг моделирования симулирует период культивирования клеток в отсутствие доксициклина и последующую рассадку клеток на 4 чашки. На этом этапе каждая N_0 клетка образует от 0 до 8 потомков. При этом, все потомки несут тот же 57 баркод, что и родительская клетка. Количество потомков определяется для каждой клетки случайно, исходя из анализа делений клеток с помощью прижизненной микроскопии. Затем клетки рассаживаются на 4 разные чашки – этот этап реализуется путем случайного присвоения каждому потомку номера в диапазоне от 1 до 4. Затем клетки на каждой чашке подвергаются случайному удалению с вероятностью от 0 до 0.9. Этот этап отражает потерю баркодов на разных стадиях культивирования клеток и приготовления библиотеки. Поскольку параметр потери материала не может быть вычислен однозначно, в модели он варьируется от 10 до 90%. Последний этап моделирования симулирует стадию репрограммирования. Из экспериментальных данных нам известно количество уникальных ДНК-баркодов, присутствующих на каждой чашке. Следовательно, мы можем рассчитать среднее количество (μ_i) и дисперсию (σ_i) этого параметра, предполагая нормальность распределения. На этом этапе моделирования создается четыре случайных числа BA1...BA4 согласно нормальному распределению с параметрами μ_i и σ_i . Затем с каждой чашки выбирается такое количество баркодов, чтобы, с одной стороны, эти числа соответствовали числам BA1...BA4, а с другой стороны, чтобы общее число выбранных баркодов (BA1 +...+ BA4) было приблизительно равно общему числу баркодов на всех чашках в эксперименте. Клетки с выбранными (т.е. репрограммированными) баркодами делятся на подгруппы так, чтобы сестринские клетки оказались в одной подгруппе. Этот шаг необходим, поскольку клетки могут нести одинаковый баркод и не быть при этом сестринскими (такие клетки будут отнесены в разные подгруппы). Затем, только одной клетке из каждой подгруппы присваивается возможность быть репрограммированной. В этом случае вероятность передать способность к репрограммированию по наследству или коэффициент наследуемости равен 0. Мы можем варьировать этот параметр от 0 до 1. В этом случае каждой из сестринских клеток присваивается заданная вероятность быть репрограммированной. Результатом моделирования является количество пересечений или общих баркодов между выбранными баркодами на разных уровнях наследуемости. Пересечения при этом могут быть двойными (между двумя любыми группами), тройными (между тремя группами), и наконец, четверными (пересечения между всеми четырьмя группами).

Мы проводили как минимум 100 запусков моделирования на каждое значение коэффициента наследуемости и потери материала (значения варьировались от 0 до 1). Результаты симуляции сравнивались с экспериментальными результатами с помощью 58 непараметрического дисперсионного анализа (критерий Краскела-Уоллиса) при уровне значимости p

4. Полученные результаты.

1. Создана и охарактеризована библиотека ДНК-баркодированных лентивирусных векторов, позволяющая эффективно маркировать клеточную популяцию, включающую до 200 000 эмбриональных фибробластов мыши.
2. Разработана компьютерная модель, позволяющая оценить вероятность синхронного репрограммирования дочерних клеток.
3. Согласно этой модели, вероятность синхронного репрограммирования дочерних клеток составляет около 10%. Это значение превышает вероятность их случайного репрограммирования за время наблюдения, что свидетельствует в пользу передачи предрасположенности к репрограммированию при делении для первичных эмбриональных фибробластов мыши.
4. Показано, что исключение из эксперимента Thy1-негативных фибробластов (15-20%), представляющих собой менее дифференцированные клетки, не приводит к уменьшению частоты синхронного репрограммирования дочерних клеток.
5. С помощью компьютерного моделирования установлено, что быстроделющиеся клетки (оценка по количеству дочерних клеток при 2-х суточном наблюдении), при меньшем содержании в исходной популяции фибробластов, являются основным источником репрограммированных клеток.
6. Быстроделющиеся клетки являются возможными кандидатами на роль «элитарных» клеток. Присвоение им большего уровня наследуемости способности к репрограммированию в компьютерной модели удовлетворяет количеству пересечений ДНК-баркодов в эксперименте.

- Эффект от использования кластера в достижении целей работы.

Обработка данных на вычислительной машине, доступной в лаборатории, заняла бы несколько дней. На вычислительном кластере НГУ – менее одного дня.

- Перечень публикаций, содержащих результаты работы (если есть). Если имеется, указать импакт-фактор журнала (Thomson Reuters, РИНЦ,...).

A.Yunusova, V.Fishman, G.Vasiliev, N.Battulin. Deterministic versus stochastic model of reprogramming: new evidence from cellular barcoding technique. //Open Biol. 2017. T. 7 №4. С. 160311. WOS IF = 4.822

- Опционально: ваши впечатления от работы вычислительной системы и деятельности ИВЦ НГУ, а также предложения по их совершенствованию.

ИВЦ НГУ имеет мощную инфраструктуру и замечательного системного администратора, который в оперативном порядке помогает решить любые проблемы. Есть только два пожелания, которые хотелось бы увидеть реализованными в будущем. Во-первых, различия в версиях ОС и системных библиотеках между узлами сильно усложняют запуск некоторых задач (иногда даже приходится переписывать код для запуска на новых/старых узлах). Во-вторых, хотелось бы иметь доступ в интернет на сервере, пусть даже небольшой по пропускной способности канал. Дело в том, что очень многое ПО заточено на автоматическое скачивание из интернета пакетов, проверку совместимости и докачивание необходимых библиотек. Для меня, например, мучением является работа с R, и в особенности с R-BioConductor по этой причине. Пакет имеет огромное количество библиотек, активно пополняется и крайне широко используется биоинформатиками, но при установке и настройке каждого дополнения возникают проблемы из-за отсутствия доступа в интернет.