

Разработка нового метода диагностики РМЖ

Возможность определения продолжительных делеций в гене BRCA1

Ментор: Горностаева Елена
phys.lena@gmail.com



Секвенирование – это прочтение нуклеотидной последовательности ДНК. Существует несколько поколений секвенирования, самым известным является секвенирование по Сэнгеру – первое поколение. В настоящее время широко распространено также секвенирование второго поколения или секвенирование следующего поколения (the Next-Generation Sequencing, NGS) – это секвенирование на высокопроизводительных платформах, которые позволяют читать миллионы фрагментов ДНК (ридов или прочтений, длиной ~ 200 пар нуклеотидов) за один эксперимент. Уровень ошибок NGS технологий довольно высок ~1% (при секвенировании 10^6 нуклеотидов - 10 тыс из них могут быть прочитаны не верно), но за счет большого количества прочтений удается отличить реальные мутации от ошибок секвенирования.

Методом NGS можно секвенировать геном человека полностью или его значительную часть. В некоторых случаях, таких как наличие у пациента какого-либо генетического заболевания, для которого известны гены и конкретные мутации в них, нет необходимости проводить полногеномное секвенирование. Тогда имеет место таргетное секвенирование – секвенирование конкретных областей генома или отдельных генов.

В нашей лаборатории исследуют (с помощью NGS) мутации в гене BRCA1, ассоциированные с раком молочной железы. Люди с мутациями в этом гене находятся в группе риска и должны обследоваться чаще, это позволяет обнаружить болезнь на ранних стадиях и с большей вероятностью вылечить ее. Метод NGS дешевле, чем секвенирование по Сэнгеру, и хорошо находит SNP (одиночные замены в геноме).

По разным данным, от 5 до 40% мутаций в BRCA1 - это протяженные делеции (сотни-тысячи пар оснований). Делеции всегда в гетерозиготе, то есть кроме мутантного экземпляра гена есть и обычный ген без делеции, поэтому в данных секвенирования их не видно. Данные секвенирования – это набор ридов, выровненных на конкретный участок генома. Для анализа таких делеций можно было бы сравнить покрытия секвенированных участков генома в разных образцах, то есть количества ридов, выровненных на данные участки. Там, где имела бы место делеция, покрытие было бы меньше \approx в 2 раза. Однако, количество ридов "плывет" от фрагмента к фрагменту и от эксперимента к эксперименту, поэтому, по всей видимости, будет недостаточно информативно.

Для поиска этих делеций предлагается использовать следующее соображение - в районе делеции прочитывается только одна копия гена,

поэтому гетерозигот быть не может. Значит, все возможные SNP на данном участке будут в гомозиготе. В соответствии с этим, предлагается проанализировать, насколько анализ отсутствия гетерозиготности на определенных фрагментах может (или не может) свидетельствовать о наличии делеции. Скорее всего будут районы, по которым такой анализ возможен, а будут такие - по которым не возможен. Во всех случаях это будет какая-то вероятность, которую можно дополнительно прицельно проверить с помощью другого экспериментального метода, qPCR.

В качестве исходного материала можно использовать данные о встречаемости SNP из базы данных dbSNP, 1000Genomes или других подобных ресурсов. У нас имеются свои экспериментальные данные по NGS анализу ~ 400 образцов больных РМЖ. В данных образцах секвенировали полную кодирующую часть гена BRCA1. Вполне возможно, что там есть делеции. Также можно искусственно генерировать данные, чтобы проверить, что метод работает.