

GoTo.Биоинформатика

Oleg Vasilev

November 16, 2016

У каждого гена есть специальный параметр - экспрессия, то, насколько интенсивно по этому гену воспроизводится белок. Есть белки с высокой экспрессией, с низкой, а бывают такие, которые вообще не воспроизводятся, при этом экспрессия конкретного гена не только может отличаться для разных людей, но изменять своё значение со временем - например, в течении дня или по мере старения. Понять что конкретно повлияло на изменение экспрессии какого-то гена не всегда просто, в частности, одни белки могут влиять на экспрессию других генов. Наша задача как раз была посвящена исследованию такого влияния.

На входе дан ориентированный взвешенный граф, в котором вершины означают гены, ребра – потенциальную зависимость между генами, а вес на ребре - степень влияния. При этом, положительный вес означает повышение уровня экспрессии того гена, на которое указывает ребро, а отрицательный - понижение. Поскольку граф был получен не на основе чистого эксперимента, а при помощи компьютерной симуляции, там было очень много лишних ребер. К примеру, если существует зависимость $A \rightarrow B$, то довольно часто в графе встречалось ребро $B \rightarrow A$ меньшего веса. Основная трудность заключалась в том, чтобы понять какие из ребер и правда означают зависимость, и какой силы.

Было решено почистить граф. Первым шагом, нашли все циклы длинны два и выкинуто более слабое ребро (это решает проблему, описанную в предыдущем абзаце). Очень важно, что ребра недостаточно сравнивать просто так. Знак на ребре имеет физический смысл, поэтому все сравнения необходимо проводить по модулю.

Потом мы долго плакали, кололись, но продолжали есть кактусы.

В качестве итогов, мы получили список вершин-кандидатов на "управляющие" гены.