

Белостоцкий А.А.

Институт проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН), и.о. научного сотрудника, belostotsky@iitp.ru

Биологические базы знаний и их применения

КЛЮЧЕВЫЕ СЛОВА:

Базы данных, базы знаний, биология, ДНК, РНК, белок, транскрипция.

АННОТАЦИЯ:

В настоящее время без применения баз данных (БД) невозможно осуществлять никакую научную деятельность. Часто прорыв в понимании наступает не в связи с кропотливой работой, с аккуратными расчетами, а просто в связи с сопоставлением различных имеющихся данных. Чаще бывает и так, что это не приводит ни к какому прорыву, однако позволяет удобно представить данные. Как общие полезные примеры, так и примеры из личного опыта, когда решения лежат на поверхности, и представлены в нижеследующем докладе.

Пример о вульгарном удобстве состоит в том, что можно совместить данные о белок-белковом взаимодействии [1] и геномную информацию [2] (разметка регуляторных участков на ДНК) так чтобы они давали принципиально новую информацию. Мы знаем, где связываются отдельные белки на ДНК. Знаем, что есть пары белков, для которых было показано, что они взаимодействуют в тех или иных условиях. Можно предположить, что когда эти белки связываются рядом друг с другом на ДНК, то они взаимодействуют и физически в этом случае. В случае множества пар, пересекающихся по одному из белков, можно реконструировать белковый комплекс, который связывается в наперед заданном регуляторном участке генома. При этом в этот комплекс могут входить белки с разными ролями в процессе функционирования ДНК. В случае регуляции инициации транскрипции был реализован простой подход, использующий хорошо подтвержденную информацию об участках ДНК, связанных с белком (учитывались пересечения этих участков), об открытых участках хроматина и о подтвержденных парах взаимодействующих белков. Знать возможные белковые комплексы в конкретном месте генома очень существенно, например, для разработки лекарств. Адрес программы www.chip-ppi.com.

Другой пример о том, что можно анализировать не только хорошо подтвержденные данные (взяты из [2]) о связывании белка с ДНК, но и более слабые сигналы, и попросту искать корреляции этих сигналов в случае разных белков. Причем оказалось, что белки, регулирующие посадку

РНК-полимеразы III на начало читаемых ею генов, сильно «шумят» и за пределами генов, читаемых РНК-полимеразой III, однако их сигнал никак не коррелирован с сигналом в генах, читаемых РНК-полимеразой II. А интрига в том, что такие же «слабые» сигналы белков, регулирующих посадку РНК-полимеразы II на начало читаемых ею генов, положительно коррелированы со слабым сигналом РНК-полимеразой II внутри читаемых ею генов. Причем такая корреляция наблюдается в случае почти всех белков, регулирующих посадку РНК-полимеразы II на начало гена (за исключением одного белка). Несущественность этих сигналов для констатации продолжительного связывания (что правильно, ведь РНК-полимеразы едут по генам, а не стоят на месте) была проверена. Исходя из этого, можно сделать вывод о том, что РНК-полимераза II находится в комплексе с белками, регулирующими ее посадку на начало соответствующих генов, что существенно меняет существующую модель и позволяет делать много громких, но обоснованных, предположений. Работа [3].

Третий пример освещает уязвимость моделей биологии (что нормально, это же наука, она постоянно развивается, пересматривает свои части), так как биологические сети (метаболические, сигнальные и генные сети) можно перерисовать без потери смысла в виде иерархической системы связанных циклов, содержащих одну (и только одну) петлю обратной связи. При этом такие циклы выступают в виде шестеренок (физику, да и мне в далеком прошлом, приходят аналогии с примитивными колебательными моделями частиц, и соответственно материи, состоящей из таких как бы «элементарных» частиц процесса). И мелкие шестеренки, крутящиеся быстро, могут крутить большие медленные шестеренки. Это грубейшая модель (самого от нее всегда тошнило и тошнит по сей день), однако она полезна для понимания возможности множественной интерпретации одних и тех же экспериментальных заключений. При этом в такой сети имеются, конечно, и «прострелы» (передача сигнала), но она должна согласовываться с основным колебательным режимом функционирования сети. Роль такого режима может быть как в устойчивости к шумам, так и в счете времени в клетке. Работа [4].

Литература

1. Использованная база данных String: <http://string-db.org/>
2. Использованная база данных UCSC Genome Browser: <http://genome.ucsc.edu/>
3. A.A. Belostotsky "Analysis of Protein-on-DNA Binding Profiles Determined with ChIP-seq Reveals Possible Interaction of Specific Transcription Factors with RNA Polymerase II during Transcription Elongation", *Biophysics (Moscow)*, 2012, Vol. 57, No. 2, P. 140
4. A.A. Belostotsky "Conception of biological networks at the molecular level as orchestrated systems of oscillators representing interconnected modular molecular clocks", *Journal of Metabolomics and Systems Biology*, 2011, Vol. 2(2), P. 15–19.