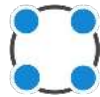


Drug synergy prediction based on transcriptomic signatures

Юрий Барбитов



Институт Биоинформатики



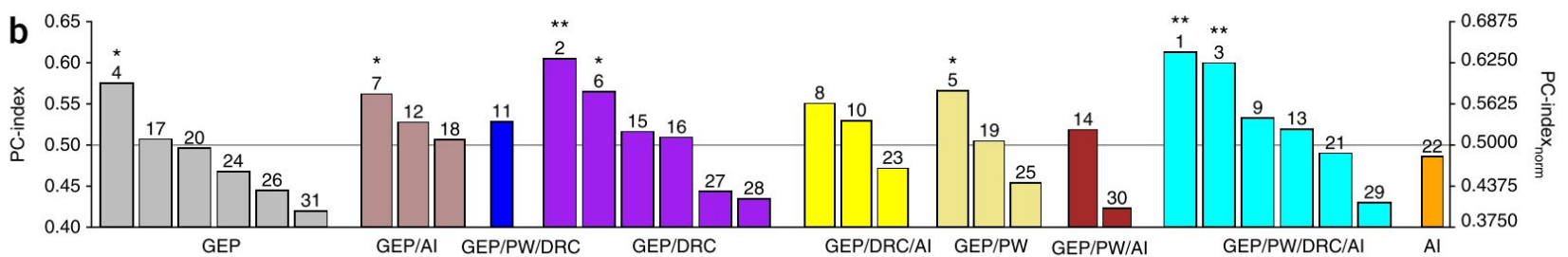
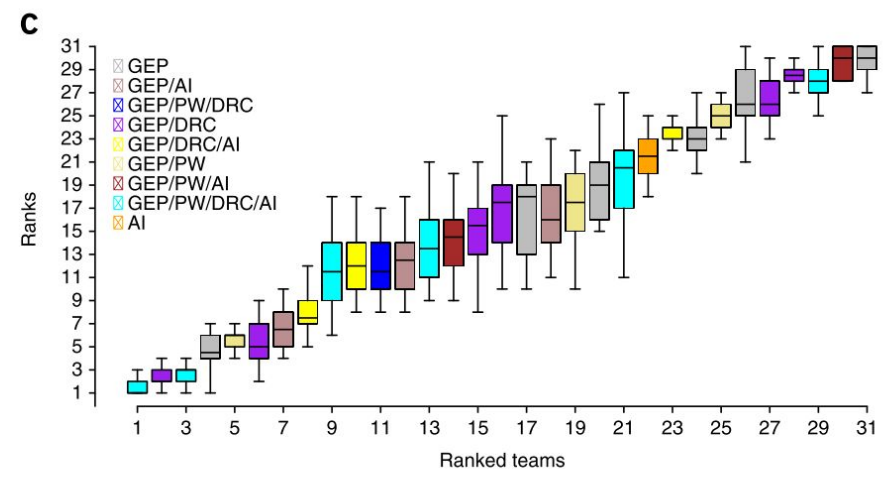
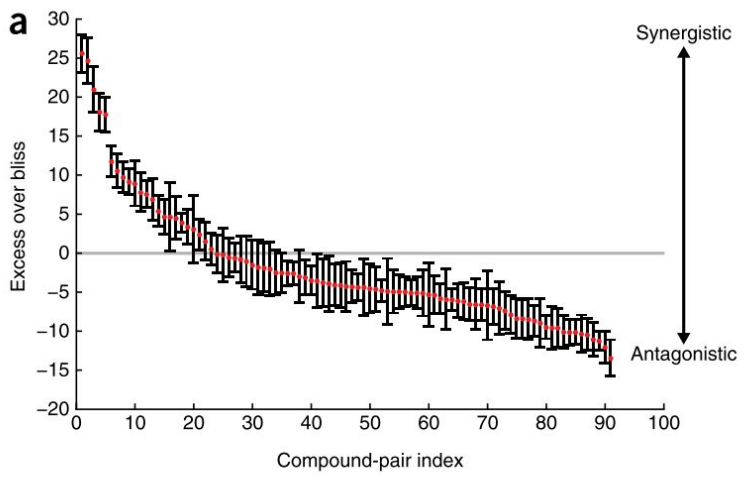
Руководитель:
Максим Артемов, Washington University in St.
Louis School of Medicine

13 сентября 2016 г.

Защита летних стажировок в Институте Биоинформатики



Drug synergy prediction challenge by NCI-DREAM

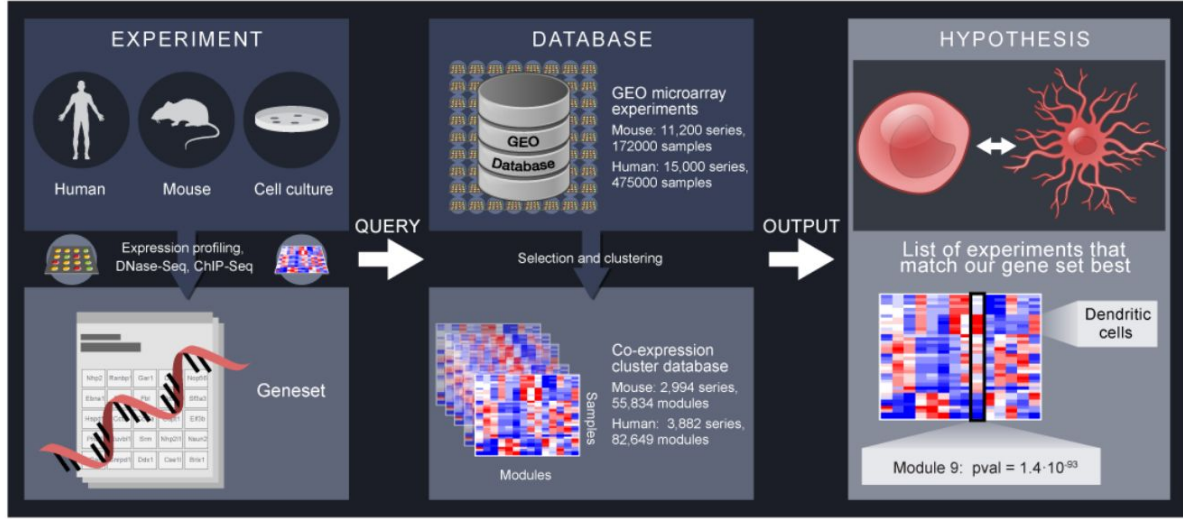


Bansal *et al.*, 2014. A community computational challenge to predict the activity of pairs of compounds. *Nature Biotechnol*



Predeus *et al.*, unpublished

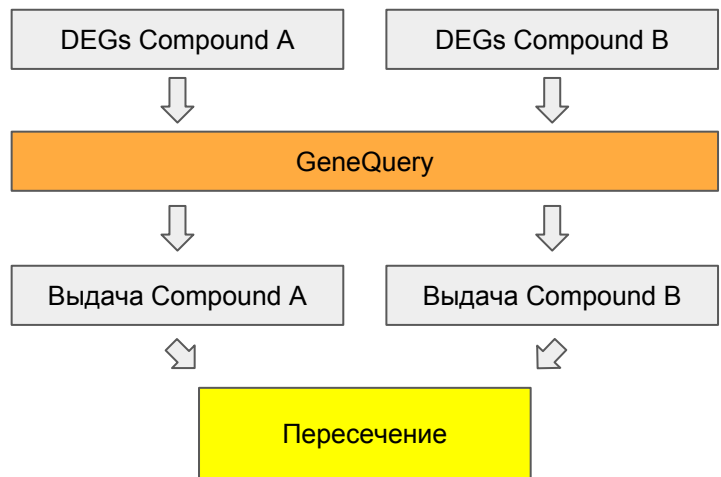
- 11491 GSE, автоматически кластеризованные при помощи WGCNA (для *Homo sapiens*)
- Точный тест Фишера с бутстрапом для поправки на случайные совпадения



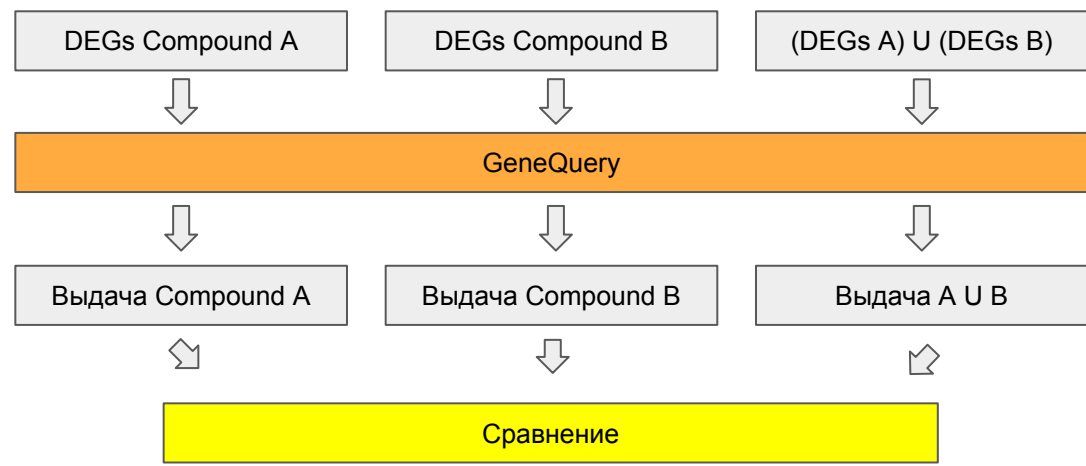


Поиск синергических транскрипционных эффектов

Подход 1



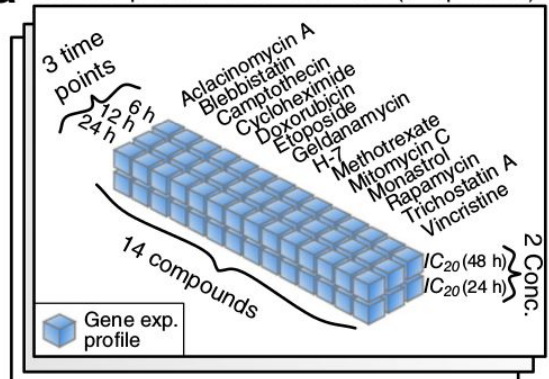
Подход 2



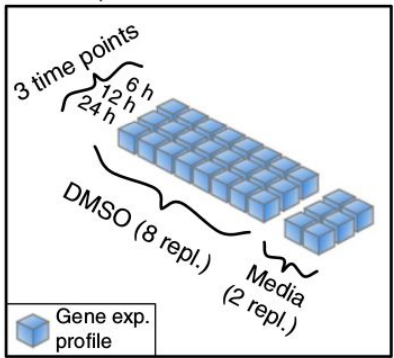


Данные NCI DREAM Challenge

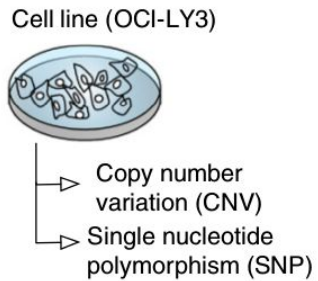
a Gene expression after treatment (3 replicates)



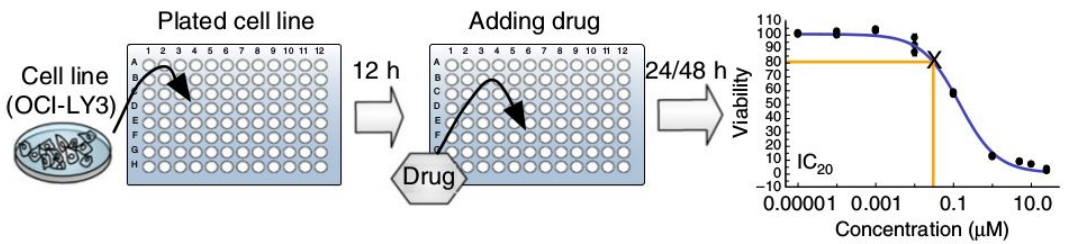
Gene expression without treatment



b



c

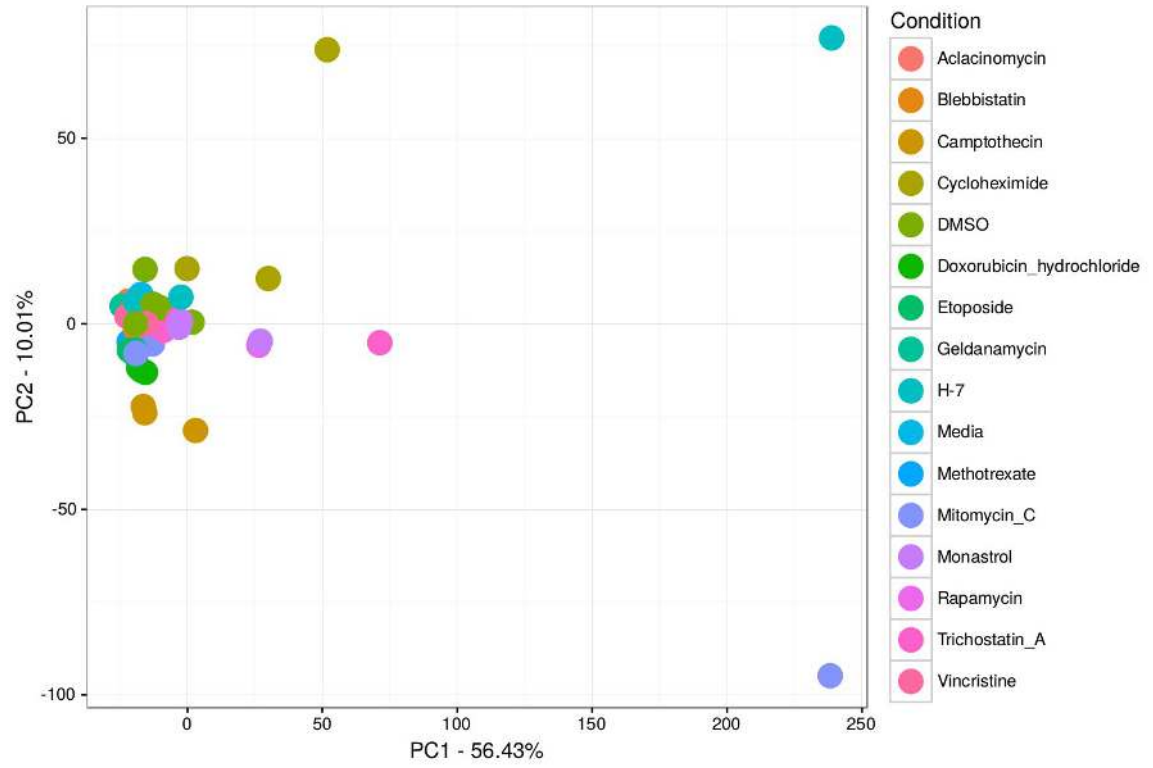
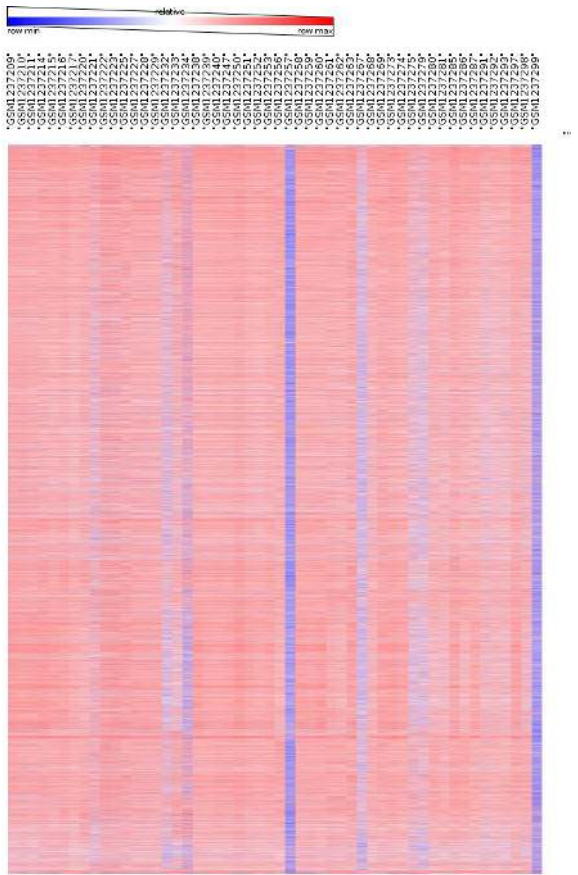


d

Compound combinations	Rank
C_1 & C_{11}	1
C_1 & C_{12}	2
...	...
...	...
C_4 & C_7	90
C_3 & C_9	91

Bansal et al., 2014. A community computational challenge to predict the activity of pairs of compounds. *Nature Biotechnol*

Данные NCI DREAM Challenge



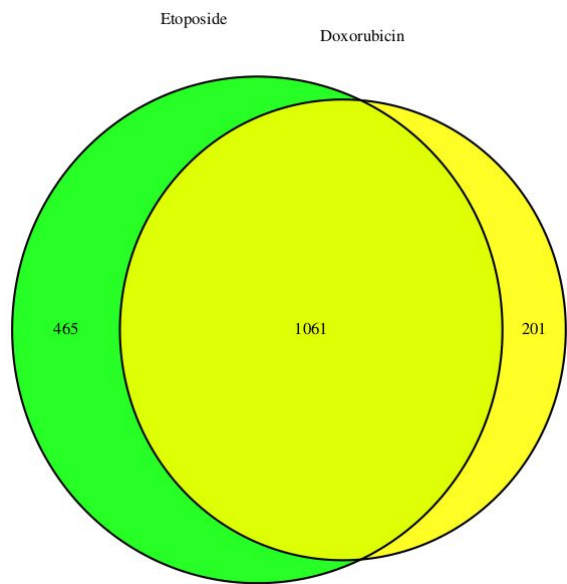
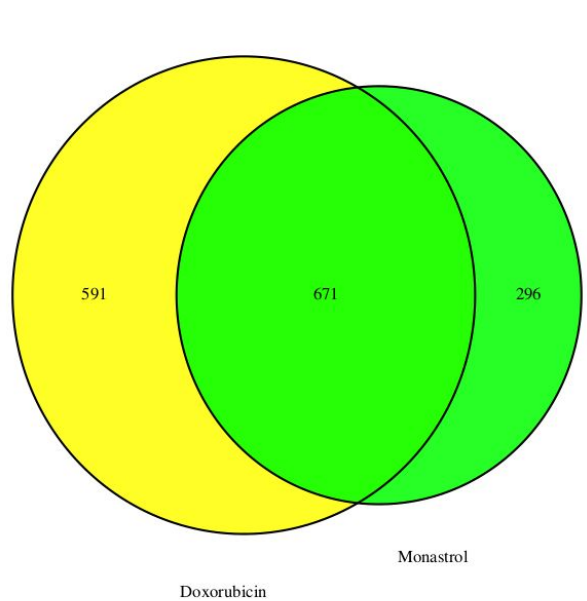
Изначально пытаемся работать с дифф. экспрессированными генами в последней временной точке и с максимальной концентрацией лекарства.



Подход номер 1 - пересечение

Пытаемся найти корреляции с размером пересечения и данными “золотого стандарта” NCI DREAM (синергический/нейтральный эффект). Например, посмотрим на две пары:

Compound A	Compound B	EOB	SE
Doxorubicin	Monastrol	13.47904955	2.34624774
Doxorubicin	Etoposide	0.160201851	2.389470246

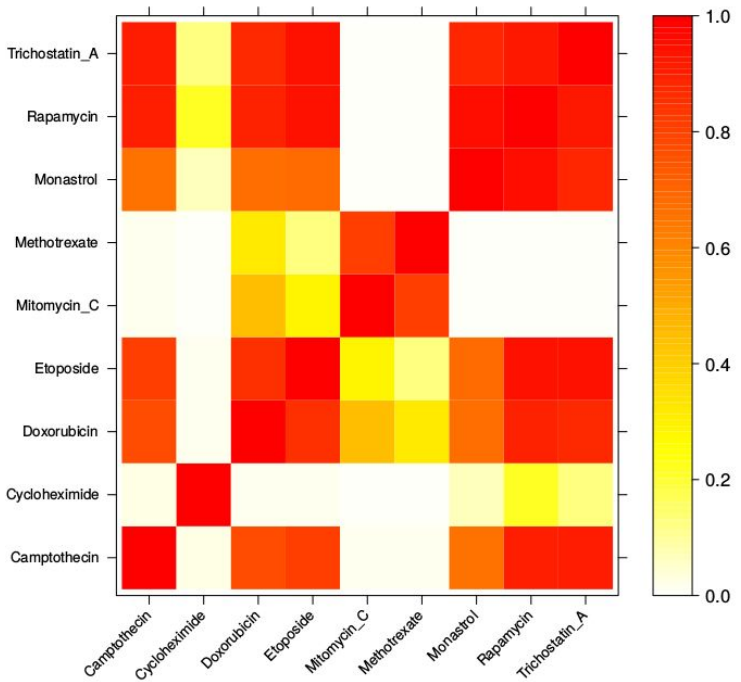
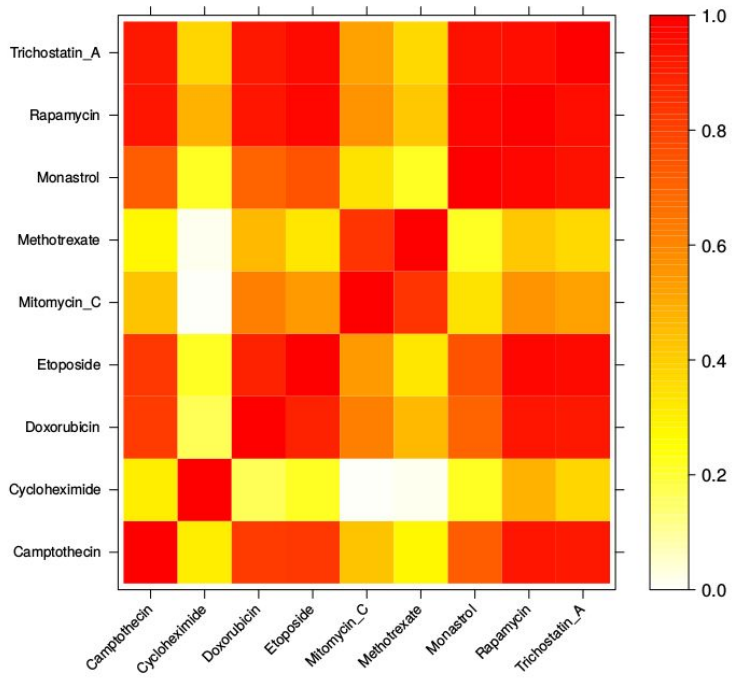




Overlap index

Считаем на двух выдачах т.н. индекс перекрывания

$$\text{Overlap_index} = |A \cap B| / (\min(|A|, |B|))$$



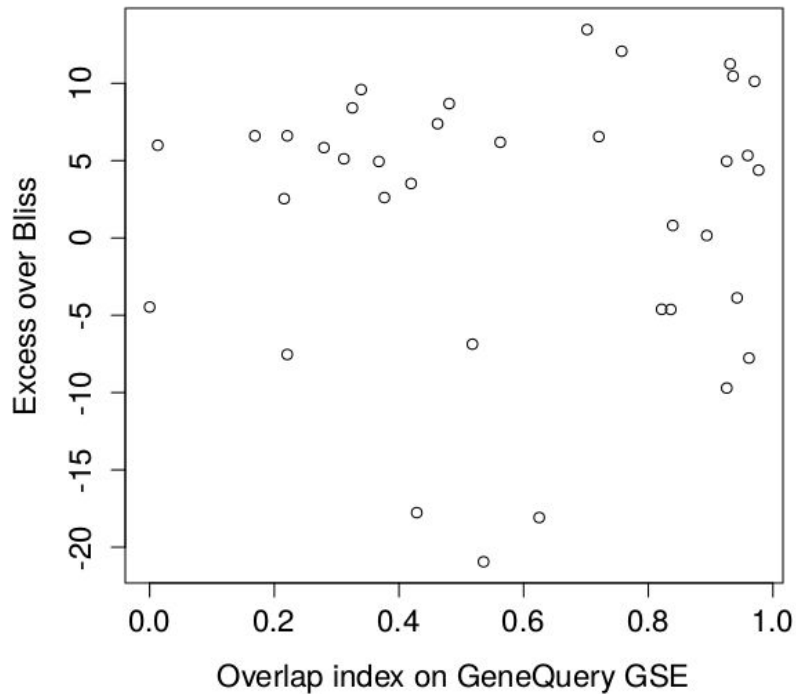
Подсчет по 9 лекарствам, имеющим интерпретируемые подписи, исходя из дифф. экспрессии в последней временной точке.

Overlap index



Считаем корреляцию с метрикой Excess Over Bliss

$$\text{cor}(\text{EOB}, \text{OI}) = -0.00585$$



Подсчет по 9 лекарствам, имеющим интерпретируемые подписи, исходя из дифф. экспрессии в последней временной точке.



Подход 2 - слитые списки

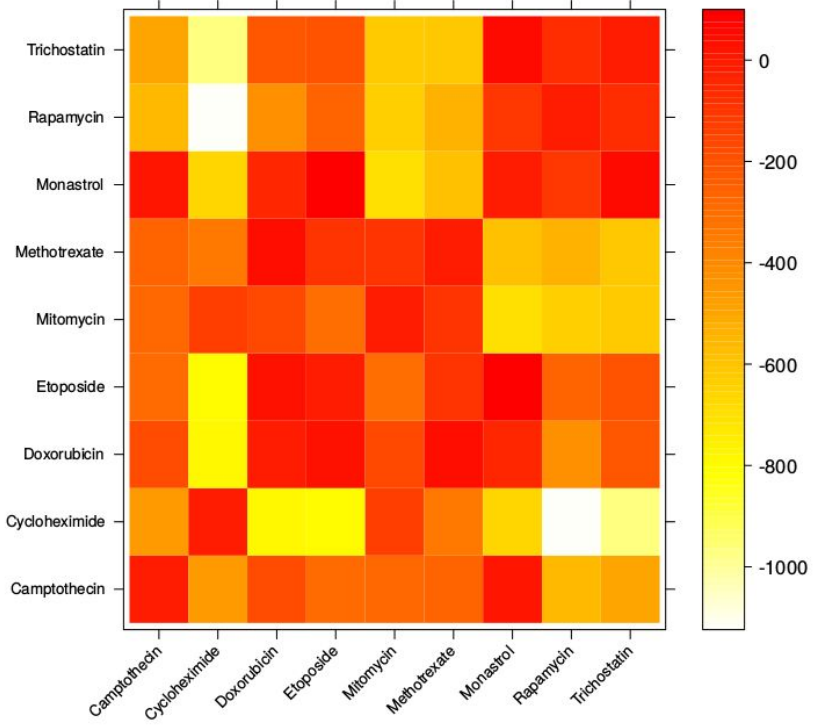
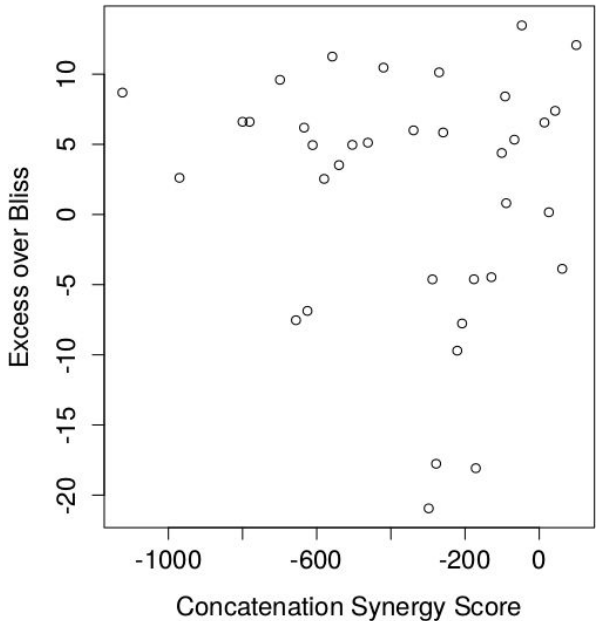
Считаем "Concatenated Synergy Score"

$$CSS = \text{genequery}(A + B) - (\text{genequery}(A) \cup \text{genequery}(B))$$

A - подпись лекарства A

B - подпись лекарства B

$$\text{cor}(\text{EOB}, \text{CSS}) = -0.128$$





Подход 2 - слитые списки

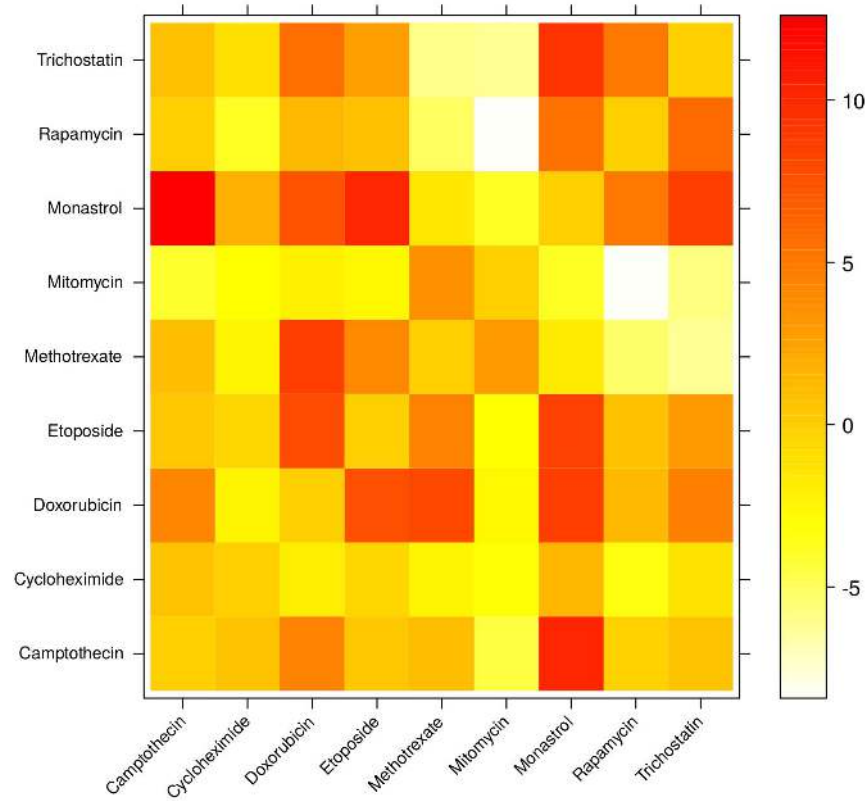
Пробуем поправлять этот скор на ожидаемое поведение. Например:

1. Берем подпись лекарства A.
2. Симулируем из множества генов GPL13667 ($|A| - |A \& B|$) случайных генов.
3. Сохраняем выдачу GQ.
4. Итерируем пп. 1-3 (100 раз)
5. Повторяем пп. 1-4 для подписи лекарства B.
6. Пересекаем выдачи лекарств A и B, дополненные рандомными генами.
7. Считаем ожидаемый размер выдачи $A + B$.
8. $Score = (|genequery(A + B)| - \text{mean}(\text{expected_size})) / \text{sd}(\text{expected_size})$

`cor(EOB, ZS) = 0.278`

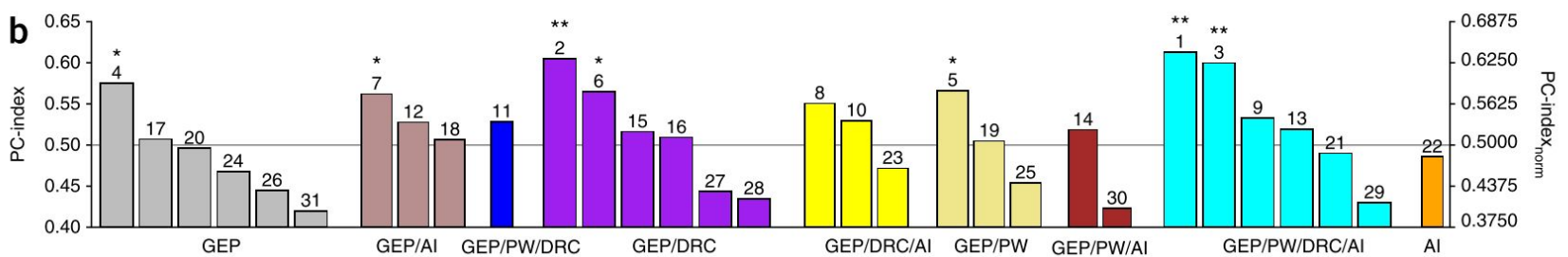
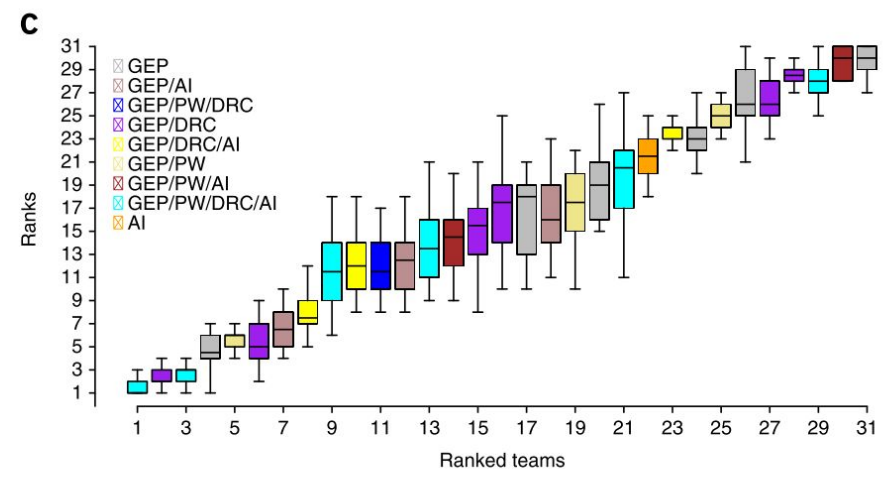
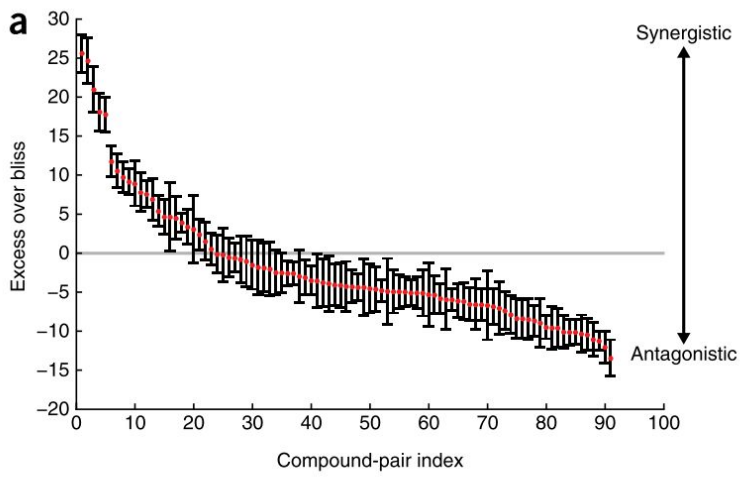
PC-index = 0.534

Мы на правильном пути!





К вопросу о рейтингах



Bansal *et al.*, 2014. A community computational challenge to predict the activity of pairs of compounds. *Nature Biotechnol*



Дальнейшие планы

На будущее:

- 1. Расширить подход на все множество генов, дифференциально экспрессированных при воздействии лекарства в любой из концентраций и временных точек.
- 2. Улучшить ноль-модель для бутстрапа.
- 3. Попробовать использовать больше потенциальной биологии для более точных предсказаний.

