



О. Н. Озолинъ

ozoline@icb.psn.ru

**Аннотация геномов по расположению
регуляторных участков**

Функциональная аннотация генома



- Поиск генов
белки
tРНК, рРНК
нетранслируемые РНК
- Определение функциональной принадлежности генов
- Реконструкция регуляторных связей в клетке

Картирование промоторов

Алгоритмы поиска промоторов:

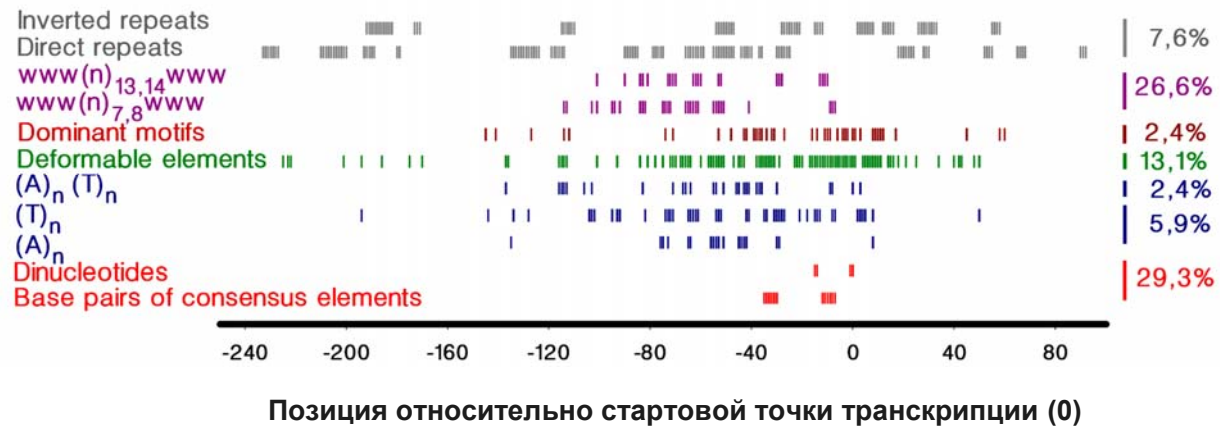


- Harr *et al.* **1983** *Nucl. Acids Res.* 11, 2943;
- Staden **1984** *Nucl. Acids Res.* 12, 505;
- Mulligan *et al.* **1984** *Nucl. Acids Res.* 12, 789;
- Harley, Reynolds **1987** *Nucl. Acids Res.* 15, 2345;
- Stormo **1990** *Methods Enzymol.* 183, 211;
- Alexandrov, Mironov **1990** *Nucl. Acids Res.* 18, 1847;
- NN** Demeler, Zhou **1991** *Nucl. Acids Res.* 19, 1593;
- NN** O'Neil **1992** *Nucl. Acids Res.* 20, 3471;
- NN** Horton, Kanehisa **1992** *Nucl. Acids Res.* 20, 4331;
- Cardon, Stormo **1992** *J. Mol. Biol.* 223, 159;
- NN** Pedersen, Engelbrecht **1995** *Intell.Syst.Mol.Biol.*, 292;
- Hertz, Stormo **1996** *Methods Enzymol.* 273, 30;
- HMM** Vanet, *et al.* **1999** *Res. Microbiol.* 150, 779;
- HMM** Yada, *et al.* **1999** *Bioinformatics* 15, 987;
- HMM** Marsan, Sagot **2000**, *J.Comput. Biol.* 7, 345;
- Leung *et al.* **2001** *Bioinformatics*, 17, 226;
- Huerta, Collado-Vides **2003** *J. Mol. Biol.*, 333, 261;
- HMM** Gordon *et al.* **2003** *Bioinformatics*, 19, 1964;
- DNA stability** Kanhere, Bansal **2005** *BMC Bioinformatics*, 6, 1;
- Gershenzon *et al* **2005** *Nucl.Acids Res.* 33, 2290;
- Down, Hubbard **2005** *Nucl. Acids Res.* 33, 1445;
- Jacques *et al.* **2006** *BMC Bioinformatics*, 7, 423;
- DNA deformability** Wang, Benham **2006** *BMC Bioinformatics*, 7, 248;
- Electrostatic properties** Sorokin *et al.* **2006** *J. Bioinf.Comput.Biol.*, 4, 455;
- Ozoline, Deev **2006**, *J. Bioinf. Comput. Biol.*, 4, 443.

Элементы, учитываемые PlatProm



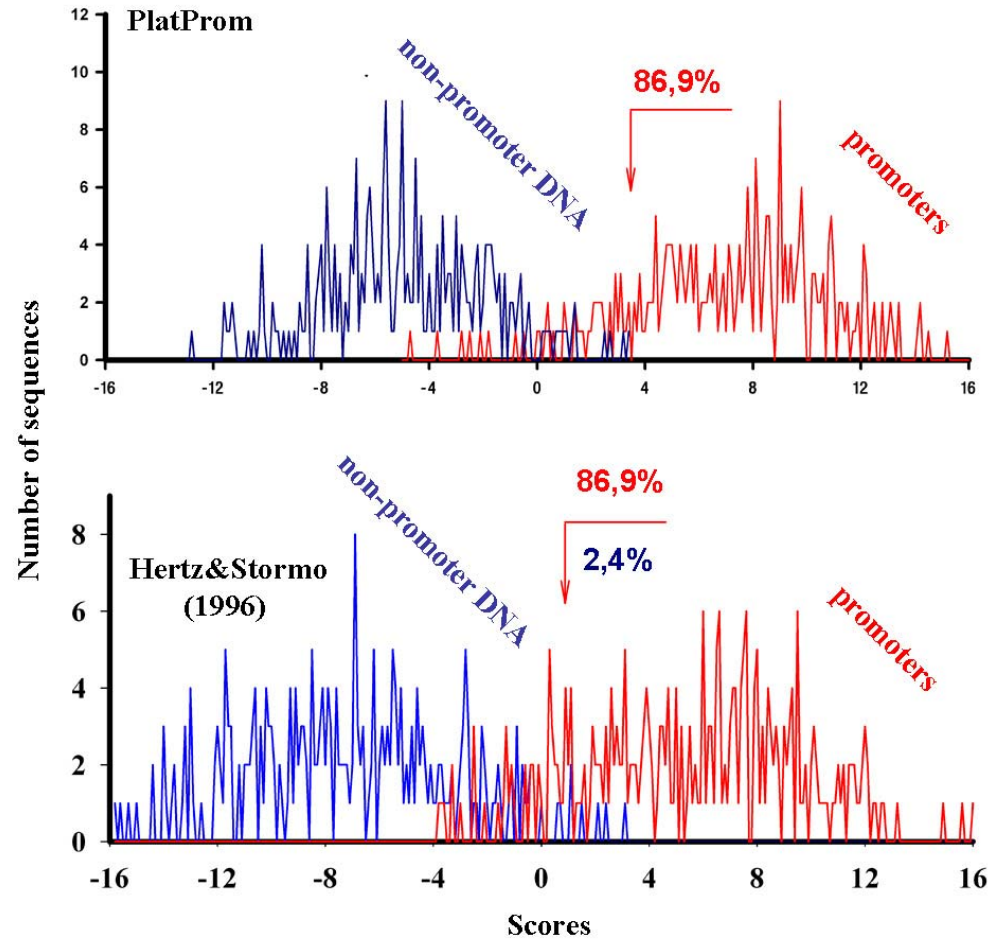
60 весовых матриц



$$S \text{ (score)} = \sum S_i$$

Platform based Promoter – serach Algorithm

"Чувствительность" PlatProm



Точность PlatProm

Точное предсказание
старта транскрипции: $(0,8 \times 0,87) \times 100 = \sim 70\%$

Обычно: <50%

PlatProm:

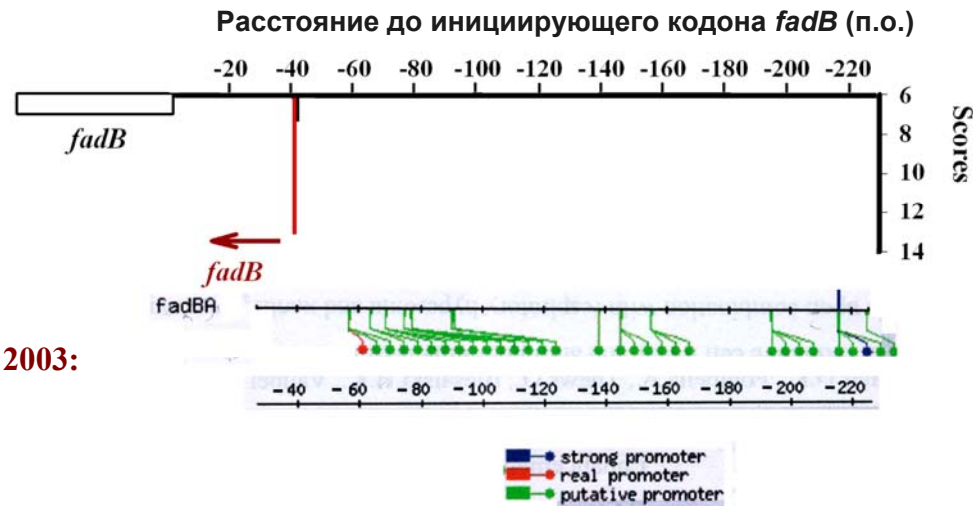
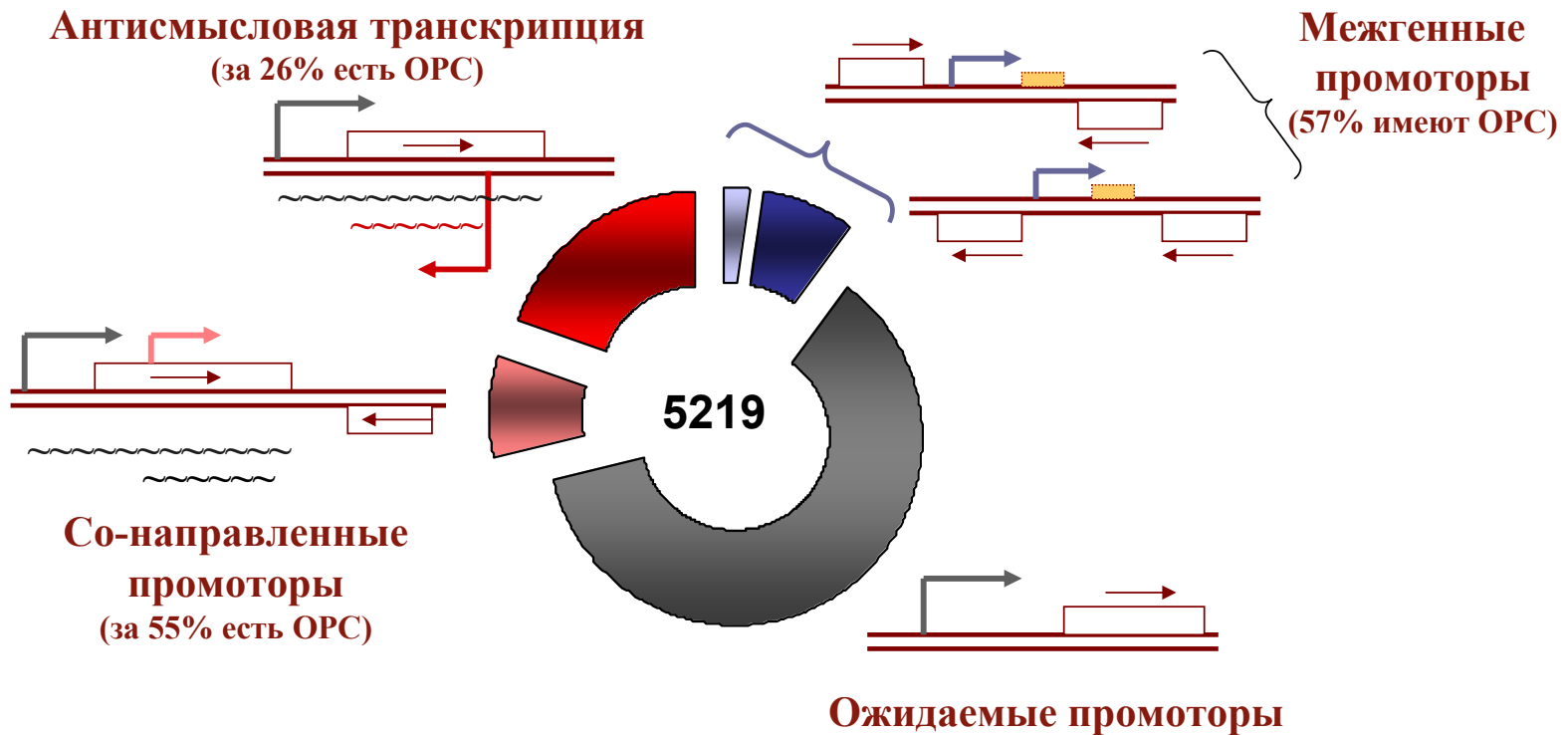


Fig. 4 in:
Huerta&Collado-Vides, 2003:

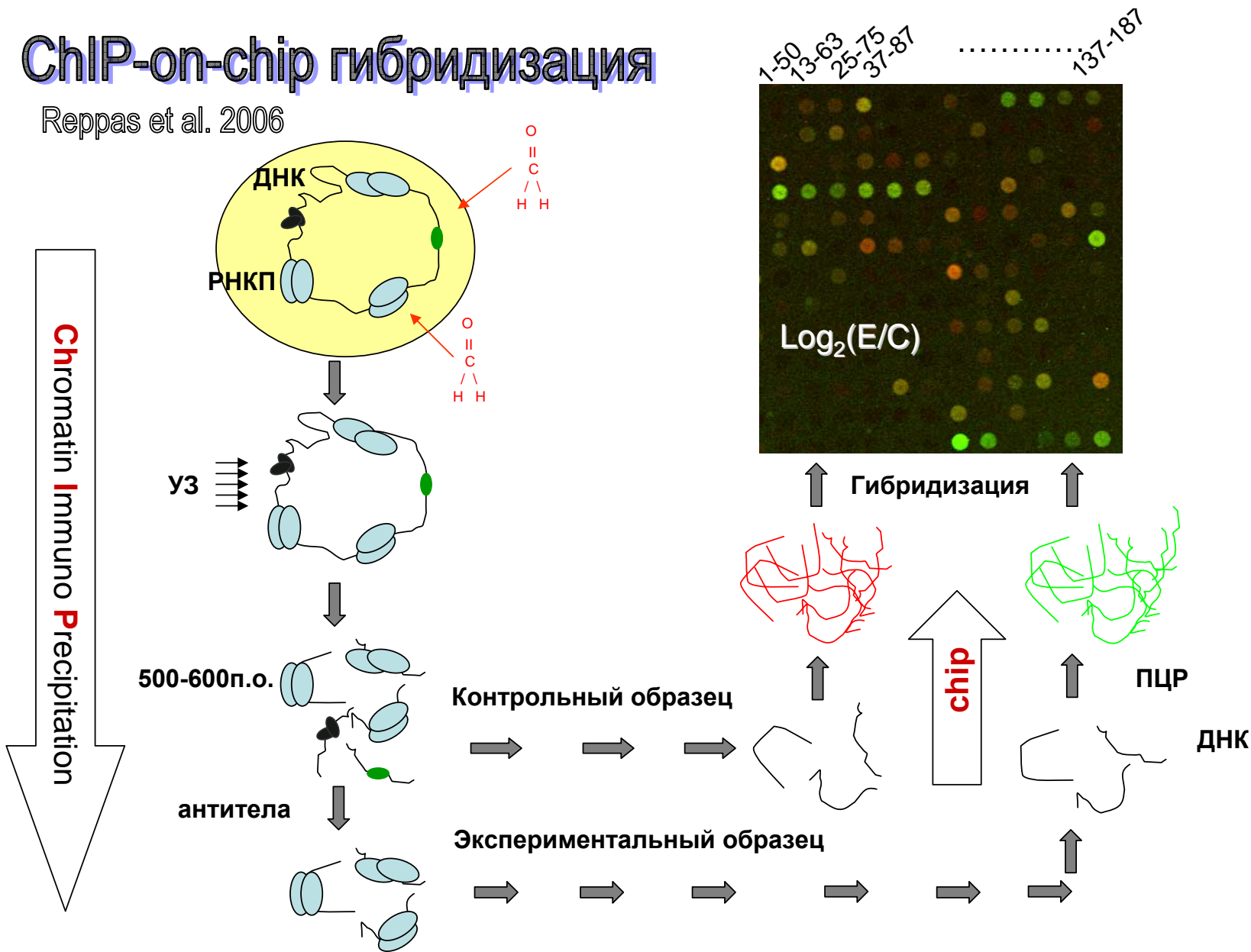
Предсказанные промоторы: распределение в геноме



1. Ошибки картирования 5'-конца гена
2. Альтернативная транскрипция
3. Непродуктивное взаимодействие с РНКП (см. постер К.Шавкунова)

ChIP-on-chip гибридикация

Reppas et al. 2006



Проверка активности предсказанных промоторов

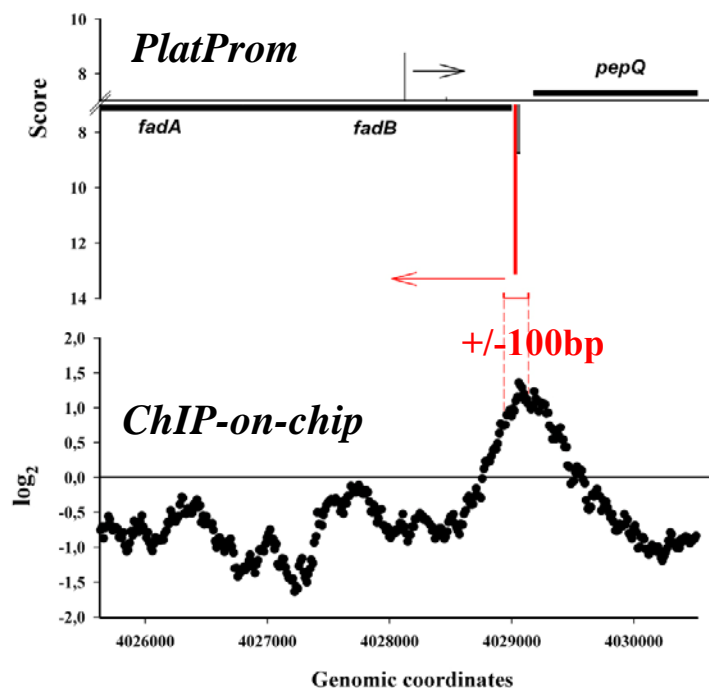
$P < 0,00004$ ($S > 7,44$)

Вопросы:

1. Сколько предсказанных промоторов взаимодействуют с РНКП?
2. Сколько мест связывания РНКП ассоциированы с промоторами?

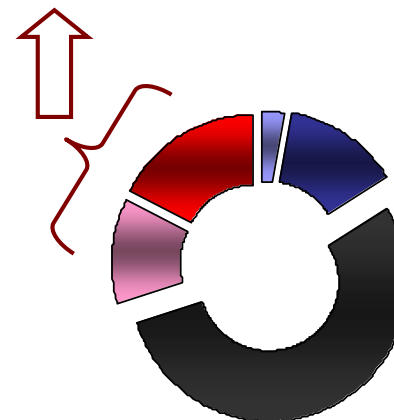


Всего точек: 30680
Ассоциированы с $\log_2 > 0$: 28815 (93,9%)
Всего кластеров: 4810
Ассоциированы с $\log_2 > 0$: 3710 (77,2%)



Вывод:

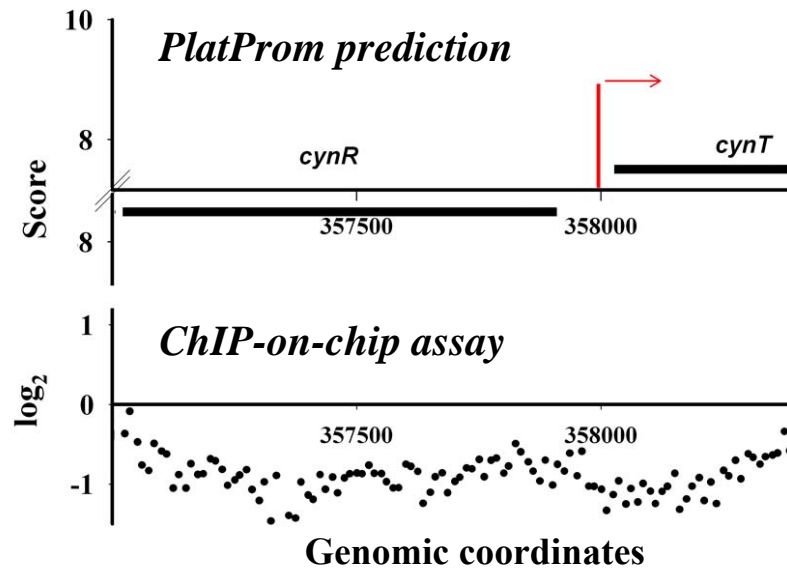
Бактериальный геном имеет связывающиеся с РНКП внутригенные промоторы



Промоторы, обнаруженные только PlatProm

Всего: 30680 точек старта

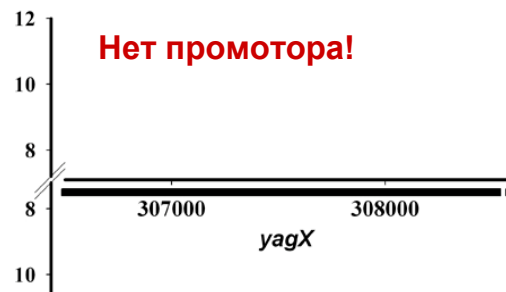
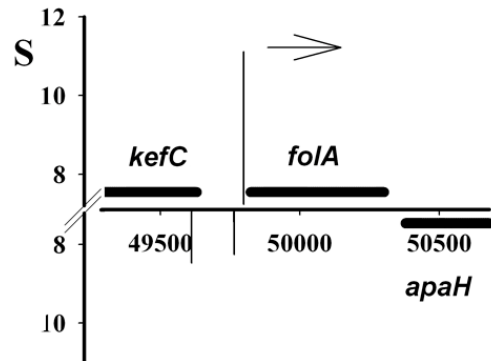
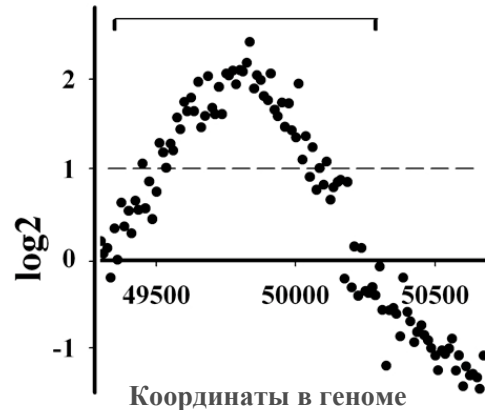
Ассоциированы с $\log_2 < 0$: 1867 (6,1%)



Вывод:

Компьютерный поиск способен предоставить более полную информацию

Ассоциированные с промотором места связывания РНКП



Всего: 51922 зондов
Ассоциированы с
 $S > 7,44$: 48418 (93,3%)

Кластеров: 1209
Ассоциированы с
 $S > 7,44$: 1049 (86,8%)

Почему внутригенные не ассоциированные с промотором комплексы с РНКП содержат σ^D ?

Заключение и перспективы



- Данные, полученные *in vivo*, в большинстве случаев подтверждают компьютерные предсказания
- Компьютерное сканирование совместно с данными ChIP-chip выявляют промоторы с необычной функцией
- Успешное картирование промоторов в геноме *E.coli* открывает перспективу использования PlatRom для анализа других геномов

Спасибо за внимание