

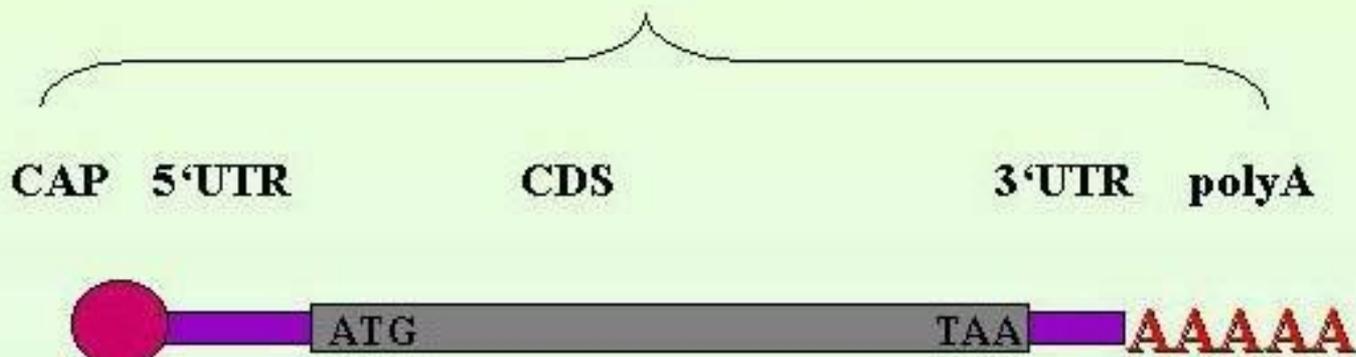


АЛЬТЕРНАТИВНЫЕ СТАРТЫ ТРАНСКРИПЦИИ И ИХ СВЯЗЬ С CpG-ОСТРОВАМИ

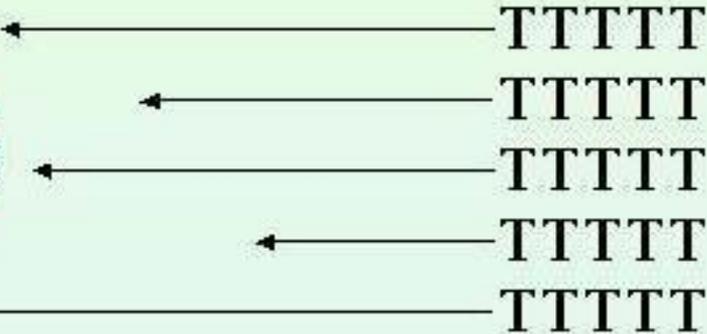
Охота за альтернативными промоторами

- Что используют? Методы клонирования кДНК только 5'-кэп-содержащих мРНК.
- Что ищут? Полноразмерные кДНК

Full-Length mRNA



Lack of Data



cDNAs from
Conventional cDNA library

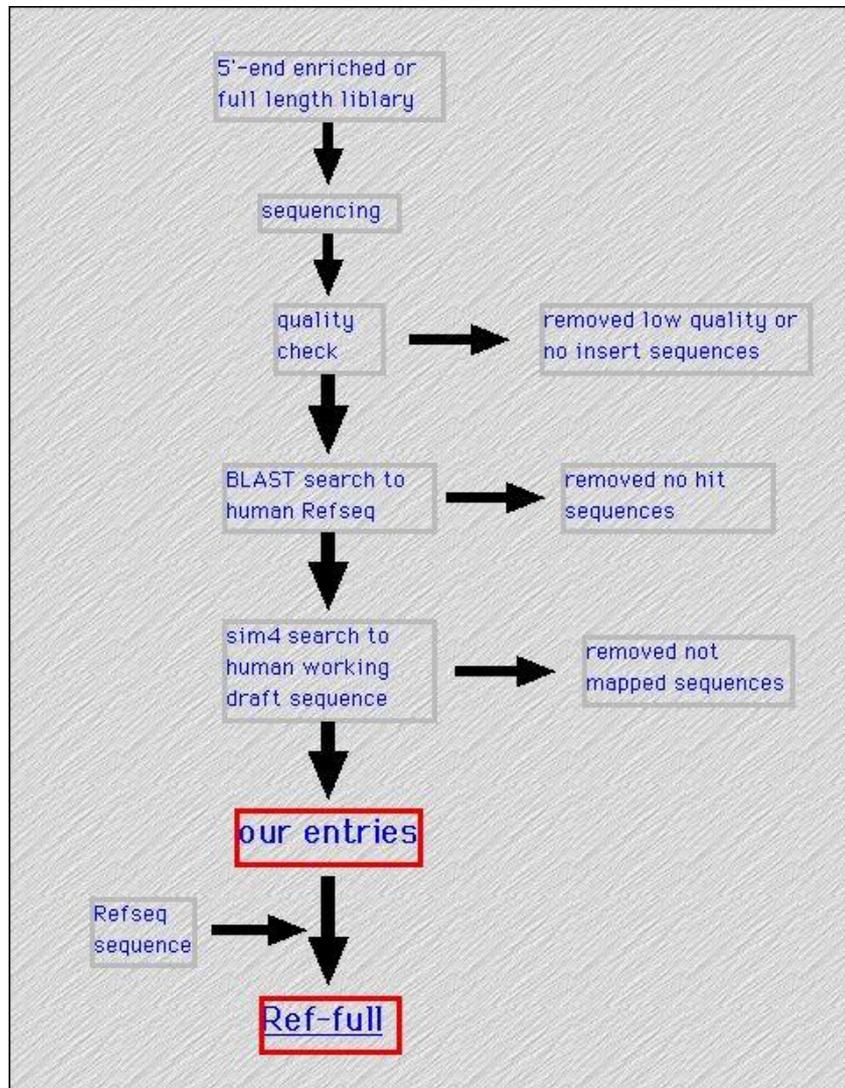
БАЗЫ ДАННЫХ

- RIKEN Mouse Encyclopedia Index (mouse full-length cDNA collection, 1,126,212 clones)
- HUNT (2001, около 9000 кДНК человека)
- NEDO human cDNA project (~1.4 млн клонов)
- ...
- **DBTSS:** База сайтов стартов транскрипции

DBTSS

	<i>#covered locus</i>	<i>#promoter</i>	<i>#total TSSs</i>	<i>#allocated clones</i>
<i>human</i>	15,262	30,964	425,117	1,359,000
<i>mouse</i>	14,162	19,023	149,876	364,487
<i>zebrafish</i>	3,061	3,382	15,198	32,263
<i>malaria</i>	1,527	N.A.	6,908	10,236
<i>schyzon</i>	3,635	N.A.	14,029	22,923

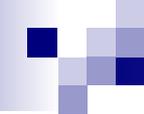
DBTSS: анализ данных



Учитывают только >95% гомологии с RefSeq

Оказалось, что более 50% RefSeq можно «удлинить» на 150 н.п. и более

«Вероятные альтернативные старты»
Должны отличаться не менее чем на 500 н.п.



	<i>human</i>	<i>mouse</i>
Total cDNAs	1,780,295	580,209
Mapped cDNAs	1,548,357	457,007
Possible truncated cDNAs	181,237	28,081
RefSeq hit cDNAs	1,359,000	364,487
* Covered RefSeq	14,628	13,704
Promoters	30,964	19,023

Какая доля генов содержит альтернативные старты (PAPs)?

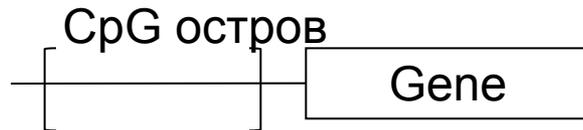
<i>#PAPs</i>	<i>#Locus</i>	<i>#included TSS positions</i>
1(PAP-less)	6954 (48%)	70175
2	3724 (26%)	67846
3	1821 (12%)	44455
4	1003 (6.9%)	32582
5	490 (3.3%)	19962
6	294 (2.0%)	13937
7	147 (1.0%)	7948
8	85 (0.6%)	4912
9	42 (0.3%)	2167
10	25 (0.2%)	1650
>10	43 (0.3%)	4140
total	14628 (100%)	269774

Альтернатива есть. И что?

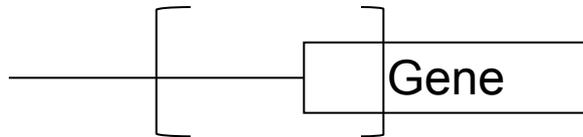
- Альтернативные старты транскрипции (PAPs) найдены для 8308 генов человека и у 4276 мышинных генов.
- Для ~1500 генов человека с PAP изоформы отличаются **только 5'-UTR**
- **Функция?**
- Для 523 ортологов человек мышь показана высокая консервативность PAPs.

СрG-острова: «маркер» старта транскрипции?

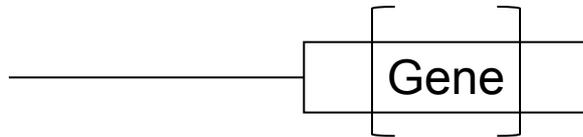
Возможное расположение СрG островов, связанных с генами



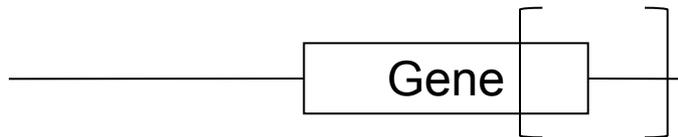
СрG в промоторной зоне



**Перекрытие с 5' концом
(~70% случаев)**



СрG внутри гена



Перекрытие с 3' концом

Какими бывают CpG острова

- **Определение Gardiner-Garden & Frommer(1987) (это определение используется в нашей работе)**
 - Длина острова больше 200 н
 - Состав G+C по меньшей мере $> 50\%$
 - Отношение наблюдаемых CpG к ожидаемым CpG : ≥ 0.6
- **Определение Takai & Jones**
 - Длина острова больше 200 н
 - G+C content: $> 55\%$
 - CpG наблюдаемое /CpG ожидаемое ≥ 0.65
 - При этих параметрах выявленные CpG острова по большей части ассоциированы с генами, и не выявляются в большинстве Alu повторов
- В человеческом геноме находятся примерно 29000 таких участков

Гены, не имеющие PAPs:

7041 генов человека и 9887 мышинных генов

Гены, имеющие PAPs:

8308 генов человека и у 4276 мышинных генов

**Нет разницы в частоте перекрытия (+-100 н.п.)
SpG-островов и стартов транскрипции (ТС)
на ген между генами с PAPs и без них
(~58% - человек, 55% - мышь)**

PAPs-гены:

- ~30% - нет перекрытия CpG-острова с ТС**
- ~5% - CpG-остров при каждом ТС**
- ~60% - CpG-остров только при одном ТС**

При этом >50% генов с PAPs содержат два альтернативных ТС

Переключение тканеспецифики?

**Для ~200 генов с PAPs с известными экспериментальными данными по экспрессии:
В 67% случаев изоформа с CpG, перекрывающим ТС – house-keeping или widely expressed**

Гены, содержащие CpG-острова, перекрывающие каждый из PAFs

- Выделена группа таких генов (источники: DBTSS, known genes, NEDO human cDNA project).
- Предварительные данные: из 79 импринтированных генов для 48 показаны PAFs, из них для 37 CpG-острова перекрывают каждый TC