Transcription of ENCODE Pseudogenes -- RACE

ENCODE Pseudogene Team
Prepared by Deyou Zheng
02-17-06
5’-RACE Experiment Design

- **Pool A (51)**
  - Primers Specific for pseudogene

- **Pool B (23)**
  - Primers with 0-3 mismatch

- **Pool C (48)**
  - Primers with 0 mismatch

- **Pool D (35)**
  - Primers with 1-3 mismatch

- **49 Non-Processed (NPS)**
  - 23

- **108 Processed (PS)**
  - 26
  - 25
  - 48
  - 35

Pseudogene

Parent gene

5’ RACE Primer
420 RACE Frags

Pool B (23)
- Primers with 0-3 mismatch
- 100 pool B specific / 196 in pool B
- 6 pseudogenes (6 NPS)

Pool A (51)
- Primers Specific for pseudogene
- 28 pool A specific / 90 in pool A
- 9 pseudogenes (4 NPS, 5 PS)

Pool C (48)
- Primers with 0 mismatch
- 138 pool C specific / 244 in pool C
- 26 pseudogenes (26 PS)

Pool D (35)
- Primers with 1-3 mismatch
- 34 pool D specific / 116 in pool D
- 7 pseudogenes (7 PS)

Pseudogene

Expected 5' RACE Primer Location

All pools 48 Pseudogenes have RACEFrgs (10 NPS, 38 PS).
420 RACE Frags

**Pool B (23)**  
Primers with 0-3 mismatch  
100 pool B specific RACE Frags

**Pool A (51)**  
Primers Specific for pseudogene  
28 pool A specific RACE Frags

**Pool C (48)**  
Primers with 0 mismatch  
138 pool C specific RACE Frags

**Pool D (35)**  
Primers with 1-3 mismatch  
34 pool D specific RACE Frags

1 pseudogene only

2 parent gene only

6 both

3 parent gene only

2 parent gene only

3 both

---

When pseudogenes and parent genes are within ENCODE regions:

1 -- pseudogenes transcribed, but not RACEFrag from the parent genes.
7 -- parent genes transcribed, but not RACEFrag from the pseudogenes.
9 -- both pseudogenes and their parents are transcribed.
Examples of Probe Intensities from RACE-Chip
Pool A (51) Primers Specific for pseudogene

Pool C (48) Primers with 0 mismatch

Pool B (23) Primers with 0-3 mismatch

Pool D (35) Primers with 1-3 mismatch

100 pool B specific RACE Frags

28 pool A specific RACE Frags

138 pool C specific RACE Frags

34 pool D specific RACE Frags

Use RACEFrags uniquely mapped to pseudogenes (≥50 nt sharing <85% sequence identity to other places)

2 NPS

6 (4 NPS, 2PS)

1 PS

2 PS

Conclusion:
6 / 49 non-processed pseudogenes have RACEFrag;
5 / 108 processed pseudogenes have RACEFrag.
Results in Cartoons

chr11
A  E  C  D
1 3  3 0

chr5
A  B  C  D
1 1

chr20
A  B  C  D
Pool

chr13
A  E  C  D
1 1

chr5
A  E  C  D
1 1

chr7
A  E  C  D
1 1 1 0 1 0

Pseudogene region

RACE Frag (#N in the genome it can map, #N of pieces unique to this region.)

1kb upstream
### Pseudogenes and Transcription Evidence

<table>
<thead>
<tr>
<th></th>
<th>TAR / transfrag</th>
<th>CAGE</th>
<th>DiTag</th>
<th>racefrag</th>
<th>EST / mRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAR / transfrag</td>
<td>105</td>
<td>14</td>
<td>2</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>CAGE</td>
<td></td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>DiTag</td>
<td></td>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>racefrag</td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>EST / mRNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21</td>
</tr>
</tbody>
</table>

Methods for intersecting a pseudogene with the five data sets: