Detection of New Transcripts with Oligonucleotide Arrays

How can one use the plethora of expression data?

Presented by Stefan Röpcke
Thesis - Overview

Data analysis of Affymetrix-Chips
(Normalisation, data condensation, DB)

Cancer research
- reanalysis of public raw data
- Bhattacharjee et al (PNAS 2001)
- Garber et al (PNAS 2001)
- comparison of technologies
- comparison to literature data

Interdependencies of gene structure and expression
- yeast, fly
- reanalysis of public data

Transcript detection
- analysis of expression of the same genomic locus
- evidence for antisense transcripts
- artefact or biology?
Oligo Array Experiment

Extraction of poly-A-RNA

Amplification and labeling of the RNA

Fragmentation, hybridisation and staining
Overview – Data Analysis

1. Feature level normalisation: offset substraction, division by the median

2. Data condensation:

Wilcoxon test
• nonparametric, paired test
• tests whether the PM is brighter than the MM values

Third-Quartile-Method
75% percentile of the matching oligos (PM)

Result: ONE representative expression value and ONE detection score per probeset and chip
Implementation & Storage

CEL-files:
- an intensity and a deviation value for each feature

Data about:
- the sequences
- the experiments
- the samples

- Quality control
- Normalization

- Wilcoxon test
- expression value calculation

Perl database interface

expression database

metagen
Data analysis of Affymetrix-Chips
(Normalisation, data condensation, DB)

Cancer research
- reanalysis of public raw data
- Bhattacharjee et al (PNAS 2001)
- Garber et al (PNAS 2001)
- comparison of technologies
- comparison to literature data

Interdependencies of gene structure and expression
- yeast, fly
- reanalysis of public data

Transcript detection
- analysis of expression of the same genomic locus
- evidence for antisense transcripts
- artefact or biology?
Expression Analysis of Lung Cancer Samples
Comparison of Oligo and cDNA Array Data

Squamous cell carcinoma versus normal lung

Adenocarcinoma versus normal lung

> 0: higher in tumor, < 0: lower in tumor

X-Axis: Average LogRatio on cDNA-Arrays

Average LogRatio on Oligo-Arrays

RED points: most “differential“ genes in one of the data sets.
Thesis - Overview

Data analysis of Affymetrix-Chips
(Normalisation, data condensation, DB)

Cancer research
- reanalysis of public raw data
- Bhattacharjee et al (PNAS 2001)
- Garber et al (PNAS 2001)
- comparison of technologies
- comparison to literature data

Interdependencies of gene structure and expression
- yeast, fly
- reanalysis of public data
- avg expression vs gene length
- length of introns/exons

Transcript detection
- analysis of expression of the same genomic locus
- evidence for antisense transcripts
- artefact or biology ?
Highly expressed Genes are shorter

Yeast expression data set from Cho et al [PNAS 1998]

6200 predicted yeast genes

2820 annotated genes only (MIPS, SGD)

Length distribution of non annotated genes
Detection of Putative Transcripts

- Analysis of probesets from the same locus
- Evidence of antisense transcripts
- ? Real transcripts or technical artefact ?
Data Set

Sequence sets on metaGen chip I
• 6117 probesets (20 oligos PM-perfectly matching 25 bases + 20 MM)
• circa 4000 different genes
• 1066 Sense-antisense sequence pairs
• 588 CDS-UTR sequence pairs

Subset used for this presentation
• Blast against RefSeq database (NCBI)
• Criterion: exactly one almost complete hit (at least 18 oligos)
-> 102 Sense-antisense sequence pairs

310 Hybridisations with poly(A+)-RNA
• 250 Samples from cancer patients
• 60 Samples from cell lines
How to compare two different probesets on one chip? *(sense vs antisense probeset)*

**Data Analysis**

- **Third-Quartile-Method**
  - 75% percentile of the matching oligos (PM)

- **Wilcoxon test**
  - nonparametric, paired test
  - tests whether the PM is brighter than the MM values

- **Detection Call**
  - p-Value > 0.05: **absent**
  - p-Value < 0.05: **present**
### Counting Schema

#### Reverse complement

<table>
<thead>
<tr>
<th></th>
<th>present</th>
<th>absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>present</td>
<td>5</td>
<td>255</td>
</tr>
<tr>
<td>absent</td>
<td>3</td>
<td>47</td>
</tr>
</tbody>
</table>

#### Forward strand

<table>
<thead>
<tr>
<th>HIT ID</th>
<th>Count experiments</th>
<th>ORI</th>
<th>HIT HEADER</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_017657</td>
<td>47 3 250</td>
<td>10 +/-</td>
<td>hypothetical protein FLJ20080 (FLJ20080), mRNA</td>
</tr>
<tr>
<td>NM_022473</td>
<td>47 3 255</td>
<td>5 +/-</td>
<td>zinc finger protein 106 (ZFP106), mRNA</td>
</tr>
<tr>
<td>NM_019000</td>
<td>45 0 235</td>
<td>30 +/-</td>
<td>hypothetical protein (FLJ20152), mRNA</td>
</tr>
</tbody>
</table>

* p_a means present sense probeset and absent antisense probeset
Contradictory Chip Results to the Annotation out of 102 Pairs

RED marked entries contrary to the expectations

<table>
<thead>
<tr>
<th>HIT ID</th>
<th>Count experiments</th>
<th>ORI</th>
<th>Mean of F-R</th>
<th>HIT HEADER</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_005033</td>
<td>a_a 13 p_a 187 p_p 18</td>
<td>+/+</td>
<td>0.6110</td>
<td>polymyositis/scleroderma autoantigen 1 (75kD) (PMSCL1)</td>
</tr>
<tr>
<td>NM_014943</td>
<td>183 0 127 p_p 0</td>
<td>+/+</td>
<td>0.2265</td>
<td>KIAA0854 protein (KIAA0854)</td>
</tr>
<tr>
<td>NM_015310</td>
<td>282 5 23 p_p 0</td>
<td>+/+</td>
<td>-1.6180</td>
<td>KIAA0942 protein (KIAA0942)</td>
</tr>
<tr>
<td>NM_031431</td>
<td>300 0 10 p_p 0</td>
<td>+/+</td>
<td>1.6951</td>
<td>tethering factor SEC34 (SEC34)</td>
</tr>
<tr>
<td>NM_022497</td>
<td>136 159 0 15</td>
<td>+/-</td>
<td>3.4254</td>
<td>mitochondrial ribosomal protein S25 (MRPS25)</td>
</tr>
<tr>
<td>NM_032659</td>
<td>219 89 2 0</td>
<td>+/-</td>
<td>1.6605</td>
<td>hypothetical protein MGC11138 (MGC11138)</td>
</tr>
<tr>
<td>NM_014654</td>
<td>252 58 0 0</td>
<td>+/-</td>
<td>-0.3542</td>
<td>KIAA0468 gene product (KIAA0468)</td>
</tr>
<tr>
<td>NM_024709</td>
<td>296 14 0 0</td>
<td>+/-</td>
<td>-1.2547</td>
<td>hypothetical protein FLJ14146 (FLJ14146)</td>
</tr>
</tbody>
</table>

p_a means present sense probeset and absent antisense probeset
### Evidence for Antisense Transcripts

**RED** marked entries contrary to the expectations

<table>
<thead>
<tr>
<th>HIT ID</th>
<th>Count experiments</th>
<th>ORI</th>
<th>mean of F-R</th>
<th>HIT HEADER</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_018011</td>
<td>a_a p_p p_p_p</td>
<td>10</td>
<td>24 274</td>
<td>+/- 6,2635 hypothetical protein FLJ10154 (FLJ10154)</td>
</tr>
<tr>
<td>NM_016127</td>
<td>a_p p_p</td>
<td>3</td>
<td>34 270</td>
<td>+/- 9,8122 HSPC035 protein (LOC51669)</td>
</tr>
<tr>
<td>NM_018509</td>
<td>a_p p_p</td>
<td>9</td>
<td>32 266</td>
<td>+/- 9,4732 hypothetical protein PRO1855 (PRO1855)</td>
</tr>
<tr>
<td>NM_006526</td>
<td>a_p p_p</td>
<td>26</td>
<td>37 228</td>
<td>+/- 1,0152 zinc finger protein 217 (ZNF217)</td>
</tr>
<tr>
<td>NM_018975</td>
<td>a_p p_p</td>
<td>4</td>
<td>97 209</td>
<td>+/- 5,0257 TRF2-interacting telomeric RAP1 protein (RAP1)</td>
</tr>
<tr>
<td>NM_016629</td>
<td>a_p p_p</td>
<td>19</td>
<td>98 191</td>
<td>+/- 7,6667 hypothetical protein (LOC51323)</td>
</tr>
<tr>
<td>NM_024026</td>
<td>a_p p_p</td>
<td>35</td>
<td>65 178</td>
<td>+/- 1,4489 mitochondrial ribosomal protein 63 (MRP63)</td>
</tr>
<tr>
<td>NM_016617</td>
<td>a_p p_p</td>
<td>8</td>
<td>125 175</td>
<td>+/- 3,4644 hypothetical protein (BM-002)</td>
</tr>
<tr>
<td>NM_021238</td>
<td>a_p p_p</td>
<td>5</td>
<td>146 158</td>
<td>+/- 6,7714 TERA protein (TERA)</td>
</tr>
<tr>
<td>NM_030912</td>
<td>a_p p_p</td>
<td>7</td>
<td>147 155</td>
<td>+/- 6,247 tripartite motif protein TRIM8 (TRIM8)</td>
</tr>
<tr>
<td>NM_022349</td>
<td>a_p p_p</td>
<td>38</td>
<td>81 150</td>
<td>+/- 2,8373 CD20-like precusor (LOC64166)</td>
</tr>
<tr>
<td>NM_006283</td>
<td>a_p p_p</td>
<td>14</td>
<td>156 138</td>
<td>+/- 9,6877 transforming, acidic coiled-coil containing prot, 1 (TACC1)</td>
</tr>
<tr>
<td>NM_015385</td>
<td>a_p p_p</td>
<td>101</td>
<td>43 132</td>
<td>+/- 1,7782 SH3-domain protein 5 (ponsin) (SH3D5)</td>
</tr>
<tr>
<td>NM_014050</td>
<td>a_p p_p</td>
<td>67</td>
<td>64 124</td>
<td>+/- 2,373 mitochondrial ribosomal protein L42 (MRPL42)</td>
</tr>
<tr>
<td>NM_023037</td>
<td>a_p p_p</td>
<td>86</td>
<td>91 91</td>
<td>+/- 1,558 putative gene product (13CDNA73)</td>
</tr>
<tr>
<td>NM_022781</td>
<td>a_p p_p</td>
<td>25</td>
<td>50 78</td>
<td>+/- 1,6658 hypothetical protein FLJ21343 (FLJ21343)</td>
</tr>
<tr>
<td>NM_007106</td>
<td>a_p p_p</td>
<td>36</td>
<td>78 78</td>
<td>+/- 2,0472 ubiquitin-like 3 (UBL3)</td>
</tr>
<tr>
<td>NM_016052</td>
<td>a_p p_p</td>
<td>1</td>
<td>7 77</td>
<td>+/- -1,7541 CGI-115 protein (LOC51018)</td>
</tr>
<tr>
<td>NM_032323</td>
<td>a_p p_p</td>
<td>282</td>
<td>13 15</td>
<td>+/- 2,2953 hypothetical protein MGC13102 (MGC13102)</td>
</tr>
<tr>
<td>NM_020188</td>
<td>a_p p_p</td>
<td>7</td>
<td>2 162</td>
<td>+/- -1,0526 DC13 protein (DC13)</td>
</tr>
<tr>
<td>NM_015070</td>
<td>a_p p_p</td>
<td>39</td>
<td>93 147</td>
<td>+/- 0,6826 KIAA0853 protein (KIAA0853)</td>
</tr>
<tr>
<td>NM_017689</td>
<td>a_p p_p</td>
<td>62</td>
<td>16 111</td>
<td>+/- -7,3946 hypothetical protein FLJ20151 (FLJ20151)</td>
</tr>
</tbody>
</table>

**p_a** means **present** sense probeset and **absent** antisense probeset

22 out of 102
Correlation of Expression Values of Sense and Antisense Probeset

**Few Cases**
similarly expressed annotated and antisense strand

**Most Cases**
higher and more variable expressed annotated strand
Biological Phenomenon or Technical Artefact

Sense-Antisense sequence pairs (total 102)

1. Reflection of the annotation in most pairs
2. **BUT** Contradictory chip results in 29 pairs:
   - 21 pairs of probesets (sense-antisense) are together present in more than 76 out of 310 experiments.
   - Further 8 pairs: contradictory chip results to the annotated orientation

In situ hybridization: Sense probe as standard control

bt12 Antisense (14885/94)  bt12 Sense (14885/94)
No Systematic Effect of the General Quality of Experiments

Table of Indicators for the Extremes

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Counts of</th>
<th>Background</th>
<th>Range of values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negative probe pairs</td>
<td>Saturated</td>
<td>Present calls</td>
</tr>
<tr>
<td>No_69_1_3</td>
<td>41804</td>
<td>1546</td>
<td>2394</td>
</tr>
<tr>
<td>No_138_1_1</td>
<td>41927</td>
<td>1928</td>
<td>2266</td>
</tr>
<tr>
<td>No_184_1_1</td>
<td>44132</td>
<td>328</td>
<td>2217</td>
</tr>
<tr>
<td>No_185_1_1</td>
<td>44918</td>
<td>330</td>
<td>2120</td>
</tr>
<tr>
<td>No_242_1_1</td>
<td>42214</td>
<td>1043</td>
<td>2389</td>
</tr>
<tr>
<td>No_254_1_1</td>
<td>41582</td>
<td>1202</td>
<td>2487</td>
</tr>
<tr>
<td>mdamb231_p5</td>
<td>40594</td>
<td>4449</td>
<td>2480</td>
</tr>
<tr>
<td>B61_1_1</td>
<td>41547</td>
<td>1478</td>
<td>2469</td>
</tr>
<tr>
<td>No7_3rd_15ug</td>
<td>53027</td>
<td>2</td>
<td>1297</td>
</tr>
<tr>
<td>No16_3rd_15ug</td>
<td>55847</td>
<td>4</td>
<td>983</td>
</tr>
<tr>
<td>No_27_1_1</td>
<td>50842</td>
<td>530</td>
<td>1526</td>
</tr>
<tr>
<td>No_78_1_1</td>
<td>55829</td>
<td>190</td>
<td>776</td>
</tr>
<tr>
<td>no_203x_1_2</td>
<td>50574</td>
<td>1</td>
<td>1526</td>
</tr>
<tr>
<td>C32_1_1</td>
<td>51371</td>
<td>124</td>
<td>1414</td>
</tr>
<tr>
<td>Cal6_14rere</td>
<td>51750</td>
<td>32</td>
<td>1433</td>
</tr>
<tr>
<td>Cal6_14___06_10</td>
<td>52238</td>
<td>20</td>
<td>1326</td>
</tr>
<tr>
<td>Cal6_7rere_06_10</td>
<td>51885</td>
<td>1</td>
<td>1346</td>
</tr>
</tbody>
</table>
No Systematic Effect of GC-Content

Number of samples (310) in which sense and antisense probes detect a signal.
No Systematic Effect of Preamplification

Total 247

Portion of TISSUE samples in which sense and antisense probes detect a signal (preamplified)

Total 63

Each red star represents one RefSeq transcript

Portion of Cell Line samples in which sense and antisense probes detect a signal
## Potential Binding Sites for the T7-(dT)24 Primer

<table>
<thead>
<tr>
<th>ID</th>
<th>oligo(dT)</th>
<th>end_pos</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_032323</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>NM_030912</td>
<td>TTTTCCCTTTTGGATTTTTGTTTTTGGCTTTTTGATTTTTTTTTATT</td>
<td>800</td>
</tr>
<tr>
<td>NM_024026</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>NM_023037</td>
<td>TTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT</td>
<td>600</td>
</tr>
<tr>
<td>NM_022781</td>
<td>TGGGTGGTATATTTACTGCCTTTTCTTTTTAAATTTTTTGTTT</td>
<td>1500</td>
</tr>
<tr>
<td>NM_022349</td>
<td>TTTTTTTTT</td>
<td>900</td>
</tr>
</tbody>
</table>
| NM_021238    | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
Potential Artefacts of RNA preparation

Strongest hypothesis (known in EST-data)
Result of FASTA Search
Search for U-stretches in RefSeq-mRNA (FAST-score > 25)

U-stretch 5' of the oligo set

U-stretch 3' of the oligo set

Oligoset

total 102

selected 21

remaining 81

Oligoset

U-stretch 5'

U-stretch 3'

U-stretch 5'

U-stretch 3'

57.1%

71.4%

30.9%

49.4%
Sequence Selection for Wet Experiments
Hunting for antisense transcripts

- **Sequence Selection**
  - DD3, ZNF217, Ponsin, MRPL42 - most differential
  - CGI-115, DC13, ZNF217 - F-R plot
  - FLJ10154, HSPC035, PRO1855 - highest pp
  - PRO1855, HSPC035, MRP63 - no U-stretch
  - β-Actin, GapDH (st present in 97.4%, 58.5%) - well known?
  - CGI-115 - contrary to annotation

- **Rasterfahndung**

- **OPN3/KMO** (positive control)

NO hints for systematic technical problems
DC13 – Unknown Gene
One of the best candidates

Genomische Sequenz: NT_010380.7

mRNA: NM_020188 (706 bp)

Primer: 568-545
Primer: 34-53

Exon 1
Exon 2
Exon 3
Exon 4

140
256
328
706

Genotype Sequenz: NT_010380.7

mRNA: NM_020188 (706 bp)

Primer: 34-53
Primer: 568-545

EST data: partial overlap with BM039 (unknown gene antisense)
Confirmation of the Chip Results
Hybridisation of labeled oligos on a multiple tissue northern

• Oligo-Selection from the probeset and non overlapping
  -> per transcript 7-10 Oligos (25 bases long)
• radioactive end labeling & hybridisation

3 out 4 examples where confirmed
positive control did not workc

Specificity ?
RNA In Situ Hybridisation (DC13)

Breast cancer tissue, resected and paraffin embedded

DC13 Antisense Sonde
(Sense wird detektiert.)

DC13 Sense Sonde
(Antisense wird detektiert.)

1 strong, 2 weak signals for sense and antisense out 4
Summary & Perspectives

Findings

• Antisense transcripts for a high percentage of genes
• NO systematic technical error found
• (pending) Confirmation of the transcripts by RPA

If TRUE

• In-depth analysis of EST-data
• Design of an antisense array
Acknowledgement

Dr. Mennerich (project planning, lab work)
Dr. Pilarsky (Chip Lab)
Eva Klopacki
Anke Vogel
Nicol Creutzburg (ISH)

metaGen