Microarray Gene Expression Analysis

1. Obtain the microarray dataset (samples and controls)

Sample Input File (\*.csv format)

1. Identify differentially expressed probes using GeneSpring v12.0
2. Input File (\*.csv format) – upaired T-tests, Fc>1.5, p<0.05
3. Output File

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ProbeName | p (Corr) | p | FC (abs) | Regulation | [00d](normalized) | [01d](normalized) | GeneSymbol | EntrezGeneID | GenbankAccession | RefSeqAccession |
| A\_44\_P465448 | 0.011302 | 0.002821 | 2.611741 | up | -0.20075639 | 1.1842556 |  |  | AA892298 |  |
| A\_44\_P514796 | 0.6774515 | 0.610037 | 1.06166 | down | 0.022838911 | -0.063482605 |  |  | AI232741 |  |
| A\_44\_P409518 | 0.3414002 | 0.264964 | 1.80454 | down | -0.16358726 | -1.0152187 | Gmpr | 117533 | NM\_057188 | NM\_057188 |
| A\_44\_P279262 | 0.1659328 | 0.11068 | 1.209694 | down | 0.15691741 | -0.11772442 |  |  | XM\_003750529 | XM\_003750529 |

1. Identify significant differentially expressed probes

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ProbeName | GeneID | p (Corr) | FC (abs) | Regulation | Validation |
| A\_44\_P243102 | ENSRNOG00000000007 | 0.001873 | 2.049992 | down | YES |
| A\_44\_P1013006 | ENSRNOG00000000017 | 0.001386 | 17.07008 | up | YES |
| A\_42\_P604425 | ENSRNOG00000000024 | 0.003336 | 1.607735 | up | YES |
| A\_44\_P274021 | ENSRNOG00000000029 | 0.003378 | 1.688237 | up | YES |
| A\_44\_P748832 | ENSRNOG00000000034 | 0.002324 | 3.555839 | up | YES |
| A\_44\_P309146 | ENSRNOG00000000036 | 0.027112 | 1.898563 | up | YES |
| A\_43\_P12200 | ENSRNOG00000000041 | 0.015188 | 2.044508 | down | YES |
| A\_44\_P351490 | ENSRNOG00000000047 | 0.001967 | 1.932757 | up | YES |

1. Categorize significant differentially expressed probes with different biotypes using Ensembl v77

|  |  |  |
| --- | --- | --- |
| ProbeName | GeneID | Gene Biotype |
| A\_44\_P621378 | ENSRNOG00000001226 | Protein coding |
| A\_44\_P1030298 | ENSRNOG00000001482 | lincRNA |
| A\_44\_P695277 | ENSRNOG00000001795 | pseudogene |
| A\_44\_P147123 | ENSRNOG00000002750 | lincRNA |
| A\_44\_P389123 | ENSRNOG00000002907 | Protein coding |
| A\_44\_P407948 | ENSRNOG00000003009 | lincRNA |
| A\_44\_P241439 | ENSRNOG00000003025 | processed\_pseudogene |
| A\_43\_P20568 | ENSRNOG00000003042 | pseudogene |
| A\_44\_P504374 | ENSRNOG00000003042 | Protein coding |
| A\_42\_P736889 | ENSRNOG00000003466 | antisense |
| A\_44\_P482200 | ENSRNOG00000005667 | Protein coding |
| A\_44\_P898694 | ENSRNOG00000008114 | processed\_pseudogene |
| A\_44\_P370213 | ENSRNOG00000008664 | processed\_pseudogene |

1. Identify significant differentially expressed genes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ProbeName | GeneID | p (Corr) | FC (abs) | Regulation | Validation | 1-p(Corr) |
| A\_44\_P243102 | ENSRNOG00000000007 | 0.001873 | 2.049992 | down | YES | 0.998127 |
| A\_44\_P1013006 | ENSRNOG00000000017 | 0.001386 | 17.07008 | up | YES | 0.998614 |
| A\_42\_P604425 | ENSRNOG00000000024 | 0.003336 | 1.607735 | up | YES | 0.996664 |
| A\_44\_P274021 | ENSRNOG00000000029 | 0.003378 | 1.688237 | up | YES | 0.996622 |
| A\_44\_P748832 | ENSRNOG00000000034 | 0.002324 | 3.555839 | up | YES | 0.997676 |
| A\_44\_P309146 | ENSRNOG00000000036 | 0.027112 | 1.898563 | up | YES | 0.972888 |
| A\_43\_P12200 | ENSRNOG00000000041 | 0.015188 | 2.044508 | down | YES | 0.984812 |
| A\_44\_P351490 | ENSRNOG00000000047 | 0.001967 | 1.932757 | up | YES | 0.998033 |

1. Categorize significant differentially expressed genes into protein coding genes and non-coding RNAs
2. Classify significant differentially expressed genes into long intergenic RNAs (lincRNA), pseudogene and antisense RNAs (asRNA)
3. Detect putative targets for lincRNAs

Input File

1. **Protein Coding Location File**

|  |  |  |  |
| --- | --- | --- | --- |
| GeneName | Chromosome | Start | End |
| Lrp11 | 1 | 3400376 | 3428890 |
| Pcmt1 | 2 | 3433956 | 3464919 |
| Nup43 | 9 | 3468475 | 3478295 |
| Lats1 | 9 | 3492544 | 3510720 |
| Ginm1 | 9 | 3566601 | 3582540 |
| Zc3h12d | 10 | 3635034 | 3669533 |
| Sash1 | 17 | 4470756 | 4631378 |
| Adgb | 18 | 6109841 | 6231640 |

1. **LincRNA Location File**

|  |  |  |  |
| --- | --- | --- | --- |
| GeneName | Chromosome | Start | End |
| Fam9b | 7 | 74309304 | 74321475 |
| RGD1562521 | 7 | 74801253 | 74853506 |
| Rn50\_X\_0711.1 | 9 | 71095169 | 71095806 |
| Rn50\_X\_0667.2 | 13 | 66634730 | 66635797 |
| Ct55 | 13 | 74325735 | 74338453 |
| LOC100302372 | 14 | 82693303 | 82730776 |
| Fam178b | X | 42797774 | 42820032 |
| LOC102555077 | X | 85284401 | 85285107 |
| Rn50\_7\_1163.2 | X | 116268693 | 116274701 |
| Rn50\_7\_1164.1 | X | 116340651 | 116343364 |

**Code**

#!/usr/bin/perl

@linc =`cat lincRNA.txt`;

chomp(@linc);

undef @linc\_c;

undef @linc\_gene;

$linc\_gene="x";

$c=0;

foreach(@linc) {

chop($\_);

@col=split(/\s+/,$\_);

$linc\_c[$c][0] = $col[1];

$linc\_c[$c][1] = $col[2] - 500000;

$linc\_c[$c][2] = $col[3] + 500000;

$c++;

push(@linc\_gene,$col[0]);

}

#shift(@linc\_gene);

$num=@linc\_gene;

print "$num\n";

@degs =`cat ProteinCoding3.txt`;

chomp(@degs);

undef @degs\_c;

undef @degs\_gene;

$degs\_gene="x";

$d=0;

foreach(@degs) {

chop($\_);

@col=split(/\s+/,$\_);

$degs\_c[$d][0] = $col[1];

$degs\_c[$d][1] = $col[2];

$degs\_c[$d][2] = $col[3];

$d++;

push(@degs\_gene,$col[0]);

}

#shift(@degs\_gene);

#$num1=@degs\_gene;

#print "$num1\n";

for($i=0;$i<=$c;$i++) {

print "$linc\_gene[$i]\n";

for($j=0;$j<=3891;$j++) {

if($linc\_c[$i][0] eq $degs\_c[$j][0]) {

#print "linc chr: $linc\_c[$i][0]\t DEG chr: $degs\_c[$j][0]\n";

#if($degs\_c[$j][1] >= $linc\_c[$i][1] && $degs\_c[$j][2] <= $linc\_c[$i][2]) {

# $distance1 = $degs\_c[$j][1] - $linc\_c[$i][1];

# $distance2 = $degs\_c[$j][2] - $linc\_c[$i][2];

# print "Linc Gene Name: $linc\_gene[$i]\t linc chr: $linc\_c[$i][0]\t DE Gene Name $degs\_gene[$j]\t DEG chr: $degs\_c[$j][0]\t Distance 1:$distance1\t Distance 2:$distance2 \n";

if($degs\_c[$j][1] > $linc\_c[$i][1] && $degs\_c[$j][2] > $linc\_c[$i][1]) {

$distance1 = $degs\_c[$j][1] - $linc\_c[$i][1];

print "Gene on the left: Linc Gene Name: $linc\_gene[$i]\t linc chr: $linc\_c[$i][0]\t DE Gene Name $degs\_gene[$j]\t DEG chr: $degs\_c[$j][0]\t Distance from linc start:$distance1\t \n";

}

if($degs\_c[$j][1] < $linc\_c[$i][2] && $degs\_c[$j][2] < $linc\_c[$i][2]) {

$distance2 = $linc\_c[$i][2] - $degs\_c[$j][1];

print "Gene on the right: Linc Gene Name: $linc\_gene[$i]\t linc chr: $linc\_c[$i][0]\t DE Gene Name $degs\_gene[$j]\t DEG chr: $degs\_c[$j][0]\t Distance from linc end :$distance2\t \n";

}

}

}

}

**Output File**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | lincRNA | linc chr | DEG\_PC | DEG chr |  |  |
| Gene on the right | Ino80dos | 1 | Rb1cc1 | 1 | Distance from linc end | 609632 |
| Gene on the right | 1700020I14Rik | 2 | Itga8 | 2 | Distance from linc end | 354118 |
| Gene on the right | 4731419I09Rik | 3 | Hey1 | 3 | Distance from linc end | 771424 |
| Gene on the right | RP23-61N4.3 | 6 | Gng11 | 6 | Distance from linc end | 602367 |
| Gene on the right | Gm26673 | 6 | Thsd7a | 6 | Distance from linc end | 862550 |
| Gene on the right | Gm26827 | 7 | Peg3 | 7 | Distance from linc end | 413420 |

1. Detect putative targets for pseudogene
2. Detect putative targets for asRNA

