## Data Analysis of DNA Microarrays



Can we detect knockdown of gene expression using DNA microarrays?

Starting with two biological samples



## Microarray Measurements

## Signal: Spotted arrays

Spotted microarrays


Signal is average of pixel
intensities of spot
2 numbers per spot

Red=500
Green=100
Red/Green=5 (5 Fold Greater)

## Processing microarrays: Scanning and Image analysis



File is large
Need to truncate

## Spot Intensity: Mean or Median?

- Which is more affected by extremes?
- Which is better estimate of spot intensity?


All pixels of a spot are used to calculate a Mean or Median Intensity

## Subtracting Background




碞最事。


| 2 | FEP |
| :--- | :--- |
| 3 |  |

 en＿reject＿SpotAnaly SpotAnaly SpotAnaly SpotAnaly SpotAnaly SpotAnaly SpotAnaly SpotAnaly SpotAnaly SpotAnaly SpotAnaly SpotAnaly SpotAnaly SpotAnaly OutlierFlaç OutlierFlaç O | 4 | tionOn |
| :--- | :--- |
| 5 | tialDetr |
| 6 |  |

$\frac{6}{6}$ end
endOn BGSubtrac BGSubtrac BGSubtrac BGSubtrac BGSubtrac BGSubtrac BGSubtrac BGSubtrac BGSubtrac BGSubtrac BGSubtrac BGSubtrac BGSubtrac BGSubtrac BGSubtrac BGSubtrac Dy ingleTextF FeatureEx FeatureEx FeatureEx FeatureExtractor＿OutputQCReportGraphText
 Ready







|  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $f x$ DATA |  |  |  |  |  |  |  |  |  |  |  |  |
| B | C $\quad$ D | E | F | G | H | 1 | J | K | L | M | N |  |
|  | (-)3xSLv1 NegativeControl | NegativeC( | 165 | 108.5 | 83 | 86 | 82 | 22.5 |  |  |  |  |
|  | (-)3xSLv1 NegativeControl | NegativeC( | 102 | 94 | 63 | 82 | 39 | 12 |  |  |  |  |
|  | (-)3xSLv1 NegativeControl | NegativeC( | 101 | 94 | 63 | 81 | 38 | 13 |  |  |  |  |
|  | (-)3xSLv1 NegativeControl | NegativeCl | 98 | 89 | 61 | 82 | 37 | 7 |  |  |  |  |
| \%ecial... | (-)3xSLv1 NegativeControl | NegativeC( | 160.5 | 116 | 83 | 87 | 77.5 | 29 |  |  |  |  |
|  | -)3xSLv1 NegativeControl | NegativeCr | 167 | 127 | 84 | 87 | 83 | 40 |  |  |  |  |
|  | -)3xSLv1 NegativeControl | NegativeC( | 102 | 93.5 | 63 | 83 | 39 | 10.5 |  |  |  |  |
|  | -)3xSLv1 NegativeControl | NegativeCl | 175.5 | 118.5 | 83 | 87 | 92.5 | 31.5 |  |  |  |  |
|  | -)3xSLv1 NeqativeControl | NeqativeC | 107 | 95.5 | 61 | 82 | 46 | 13.5 |  |  |  |  |
|  | +)E1A_r61E1A_r60_1 | E1A_r60_1 | 87 | 106.5 | 68 | 84 | 19 | 22.5 |  |  |  |  |
|  | +)E1A_r61E1A_r60_1 | E1A_r60_- | 84 | 87 | 68 | 85 | 16 | 2 |  |  |  |  |
|  | +)E1A_r61E1A_r60_1 | E1A_r60_ | 79 | 85 | 68 | 85 | 11 | 0 |  |  |  |  |
|  | +)E1A_r61E1A_r60_1 | E1A_r60_ | 77 | 84.5 | 66 | 83 | 11 | 1.5 |  |  |  |  |
| ıntents | +)E1A_r61E1A_r60_1 | E1A_r60_- | 84 | 83 | 67 | 83 | 17 | 0 |  |  |  |  |
|  | +)E1A_r61E1A_r60_1 | E1A_r60_- | 81 | 88 | 68 | 84 | 13 | 4 |  |  |  |  |
| Cells... ght... | +)E1A_r61E1A_r60_1 | E1A_r60_- | 83 | 84 | 67 | 83 | 16 | 1 |  |  |  |  |
|  | +)E1A_r61E1A_r60_1 | E1A_r60_ | 89.5 | 93 | 70 | 83 | 19.5 | 10 |  |  |  |  |
|  | +)E1A_r61E1A_r60_1 | E1A_r60_- | 89 | 89.5 | 72 | 84 | 17 | 5.5 |  |  |  |  |
|  | +)E1A_r61E1A_r60_1 | E1A_r60_- | 114 | 103 | 90 | 83 | 24 | 20 |  |  |  |  |
|  | +)E1A_r61E1A_r60_1 | F14 ${ }^{-1} 60^{-}$. | 87 | 845 | 69 | 82 | 18 | 25 |  |  |  |  |
| 79 (+)E1A_r61E1A_r60_1 |  |  |  |  |  |  |  |  |  |  |  |  |
| $79(+) \text { E1A } \_ \text {r6IE1A_r60_1 }$ |  |  |  |  |  |  |  |  |  |  |  |  |
| 79 (+)E1A_r61E1A_r60_1 |  |  |  |  |  |  |  |  |  |  |  |  |
| 79 (+)E1A_r61E1A_r60_1 |  |  |  |  |  |  |  |  |  |  |  |  |
| $79(+) \text { E1A_r6IE1A_r60_1 }$ |  |  |  |  |  |  |  |  |  |  |  |  |
| $79(+) E 1 A_{-}$r61E1A_r60_1 |  |  |  |  |  |  |  |  |  |  |  |  |
| 79 (+)E1A_r61E1A_r60_1 |  |  |  |  |  |  |  |  |  |  |  |  |
| 79 (+)E1A_r61E1A_r60_1 |  |  |  |  |  |  |  |  |  |  |  |  |
| 79 (+)E1A_r61E1A_r60_1 |  | E1A_r60_- | 80 | 90.5 | 63 | 82 | 17 | 8.5 |  |  |  |  |
| 79 | (+)E1A_r61E1A_r60_1 | E1A_r60_- | 110 | 97 | 65 | 82 | 45 | 15 |  |  |  |  |
| 79 | (+)E1A_r61E1A_r60_1 | E1A_r60_ | 112 | 102 | 72 | 84 | 40 | 18 |  |  |  |  |
| 79 | (+)E1A_r61E1A_r60_1 | E1A_r60_ | 89 | 102 | 65 | 81 | 24 | 21 |  |  |  |  |
| 79 | (+)E1A_r61E1A_r60_1 | E1A_r60_ | 96 | 98 | 62 | 83 | 34 | 15 |  |  |  |  |
| 791 | (+)E1A_r61E1A_r60_1 | E1A_r60_- | 156 | 120 | 82 | 85 | 74 | 35 |  |  |  |  |
| 791 | (+)E1A_r61E1A_r60_1 | E1A_r60_- | 90 | 85 | 68 | 82.5 | 22 | 2.5 |  |  |  |  |
| $\begin{array}{r}79 \\ \hline 70 \\ \hline\end{array}$ | (+)E1A_r61E1A_r60_1 | E1A_r60_- | 103 | 110 | 70 | 84 | 33 | 26 |  |  |  |  |
|  |  | C1A ${ }^{-10-1}$ | 00 | 00 | c7 | Q | 21 | 10 |  |  |  |  |



510 510 510 (+)eQC-39 EQC 510 (+)eQC-39 EQC 510 (+)eQC-39 EQC 6304 A_06_P10 ORF:Q0010
959 A_06_P10 ORF:Q0017
959 A 06 P10 ORF:Q0017
4760 A_06_P10 COX1
4760 A_06_P10 COX1
6303 A_06_P10 Al1
6303 A_06_P10 Al1
435 A_06_P10 Al2
435 A_06_P10 Al2
6302 A_06_P10 Al3
6302 A 06 P10 Al3
1066 A_06_P10 Al4

## Delete unused negative and positive control: eQCs




## Within-Slide Normalization

- Normalization balances red and green intensities.
- Imbalances can be caused by
- Different incorporation of dyes
- Different degradation of dye
- In practice, we usually need to increase the red intensity a bit to balance the green


## cy3 and cy5: Commonly used dyes





Light sensitivity: cy5 more easily degraded

## Let's begin the normalization process:





## Calculating Differences in Gene Expression

Spotted microarrays

2 numbers per spot
Red $=500$
Green=100
Red/Green=5 (5 Fold Greater in Red)
Red=100
Green=500
Red/Green $=0.2$ (5 Fold Less in Red)






## And NOW to the fun...

- How many genes were differentially expressed between your 2 samples?
- Was the expression of your gene of interest significantly changed between the two samples?...can we assess this directly




## Create scatter plot of log2 ratios (green versus red)

$\log 2$ red vs green


## Distribution of log2 ratios

- What are we expecting????
- What color would all of these spots be??


## Agilent Human 1A Oligo Microarray Kit (V2) with SurePrint Technology

## Catalog 60-mer Oligo



## Coverage you can count on

How many genes on the array?

Designed to truly represent well-known genes in the human jenome, Agilent's Human 1A Oligo microarray (V2) is comprised of 20,173 ( 60 -mer) oligonucleotide probes, which span conserved exons across the transcripts of the targeted full-length genes. These probes represent 18,716 well-characterized, full-length, human genes from RefSeq and Incyte's Foundation Database. Much of the sequence and annotation information used in this microarray product is available only through Agilent and Incyte. Virtually all of the genes and corresponding probes have been mapped to the human genome DNA backbone. These probes have been experimentally validated in a laboratory which provides the researcher with maximal confidence in the probes and prevents redundancy in gene coverage.

## Trends in Data

- How many changes do you see?
- What could these changes mean?
- How can we find out more about these genes and their functions?
- Which biological processes are upregulated, down-regulated, no change?
\$ Microsoft Excel - be109 array

§ Microsoft Excel－be109 array


|  |
| :---: |
| P1 |
| G |

I
K


1 rMedianSiçgBGMedia rBGMediar green corrected red corrected red normal red vs greelog2 red vs green
22693

| 88 |
| ---: | ---: |
| 86 |
| 84.5 |
| 84 |
| 107 |
| 84 |
| 85 |
| 82.5 |

## Convert Text to Columns Wizard－Step 1 of 3

The Text Wizard has determined that your data is Delimited．
If this is correct，choose Next，or choose the data type that best describes your data．
Original data type
Choose the file type that best describes your data：
© Delimited－Characters such as commas or tabs separate each field．
Fixed width－Fields are aligned in columns with spaces between each field．

Preview of selected data：

notated｜MolFunction＝ actin filament organization｜MolFı ubiquitin－dependent protein catal BioProces＝biological＿process unknown $\mid \mathrm{Mol}$ BioProces $s=$ biological＿process unknown $\mid \mathrm{Mol}$ BioProces $s=$ stress response＊｜MolFunction＝p

BioProces $s=$ biological＿process unknown $\mid \mathrm{Mol}$


包 integer Grid＿NumSubG SpotAnalysis＿C 2


To display this task p Office Clipboard on $t$ press Ctrl＋C twice．
$\boxed{3}$ Microsoft Excel－be109 array
：죠 File Edit View Insert Format Iools Data Window S－PLUS Help Adobe PDF
Type a question for help



## 



## 24 of 24 －Clipboa （अ）©｜त

## ［18 Paste All XCle

Click an item to paste：
园

國


图 integer
Grid＿NumSubGridR SpotAnalysis＿Cool


図


Q

To display this task pane Office Clipboard on the ： press Ctrl＋C twice．
Options－



| $\|c\|$ | Q24 |  | $f_{x}$ DNA replication initiation | ｜MolFunc |
| :---: | :---: | :---: | :---: | :---: |
|  | G | H | I | J |
| K |  |  |  |  |

1

| 1608 | B |
| :---: | :---: |
| 5772 | B |
| 5859 | B |
| 0993 | B |
| 9244 | B |
| 4185 |  |
| 7001 | B |
| 6227 | Bis |
| 0443 | B |
| 3062 | B |
| 7485 | B |
| 9349 | B |
| 6353 |  |

BioProces vacuolar acidification＊｜ M monovaleı BioProces mRNA catabolism｜MoIF molecular BioProces DNA dependent DNA re DNA－direc BioProces biological＿process unkn molecular BioProces biological＿process unkn molecular

BioProces not yet annotated｜MolFı dolichyl－p BioProces SRP－dependent，co－tran transporte BioProces response to oxidative st thiol－disul BioProces biological＿process unkn molecular BioProces osmosensory signaling enzyme a BioProces ubiquitin－dependent prot adenosinє

BioProces UDP－N－acetylglucosami UDP－N－ac BioProces biological＿process unkn molecular BioProces translational termination translation BioProces protein biosynthesis＊｜Mc structural BioProces biological＿process unkn molecular BioProces biological＿process unkn molecular

BioProces biological＿process unkn molecular BioProces biological process unkn molecular BioProces DNA replication initiatio chromatin BioProces translational elongation｜｜I translatior BioProces not yet annotated｜MolFu isoleucine BioProces mRNA splicing｜MolFunc not yet ar

BioProces not yet annotated｜MolFu protein tyI BioProces actin filament organizati structural BioProces ubiquitin－dependent prot proteason BioProces biological＿process unkn peptidyl－p BioProces biological＿process unkn molecular BioProces stress response＊｜MolFu protein ta；

BioProces biological＿process unkn molecular RinPrnres hinlncriral nraress unkn molerilar ${ }^{\vee}$

24 of 24 －Clipboard （3）© 1 a

## （ 0 Paste All

Click an item to paste：


## 娄 integer

Grid＿NumSubGridRows SpotAnalysis＿CookiePerc．．．

## （2）



## （2） 1

To display this task pane again，click Office Clipboard on the Edit menu or press Ctrl＋C twice．
Options－

## Good luck!!

