



GP3xCLI

GenePix Post-Processing
By
CSIRO Livestock Industries



Introduction

1. Each microarray hybridisation generates a large amount of data resulting from the scanning of two images (one red, one green).
2. It is both tedious and prone to human error to visually assess the quality of each spot on the microarray.
3. GP3xCLI is an automated AWK-based script to assess the quality of raw microarray data captured using the GenePix optical scanner.
4. Executed from the prompt line, GP3xCLI incorporates tools such as:
 1. A2PS → ASCII to Postscript translator
 2. GNUPLOT → interactive plotting utility
 3. PS2PDF → public domain postscript to PDF converter
5. On execution, GP3xCLI generates a 2-page PDF file.
6. Inspired by GP3 (<http://www.bch.msu.edu/~zacharet/microarray/GP3.html>), GP3xCLI is not intended to perform any real analysis (eg. Correction, filtering, normalisation).
7. Quality assessment is based on standard
 1. Summary Statistics
 2. Diagnostic plots



Summary Statistics

1. Total Number of Spots.
2. Number of Spots by Quality Flag (generated by the scanner).
3. For each channel (Red and Green), Number of Spots for which the foreground is less than the background intensity signal.
4. Optimality of Log-transformation → Amount of data left by each threshold of Mean to Median correlation.
5. Number of Repetitions by Genes
6. N, Mean, Std, Min and Max for Log-Ratio (Red/Green) and average (Red+Green) intensities



Summary Statistics

```
#####
#
# GP3xCLI
# GeneFix Processing Program by CSIRO Livestock Industries
# Bioinformatics Group
#
# Enquiries: Tony.Reverter-Gomez@csiro.au
# Copyright (c) 2003-2005 CSIRO Livestock Industries
#
#####

GPR Input:      f12.gpr
Processed on:   Tue Dec 27 22:46:25 EST 2005

===== IMAGE QUALITY =====
Total No. of Spots -----> 19200

QUALITY FLAG   SPOTS
-----
0              14467
-100           12
-50            4721

Red dye with Background >= Foreground ---> 892
Green dye with Background >= Foreground ---> 915

Median to Mean Correlation Analysis:

          DATA LEFT
          RED        GREEN
Corr     Raw  Log2  Raw  Log2
-----
> 0.00  19200 19200  19200 19200
> 0.20  19199 19200  19199 19200
> 0.40  19193 19200  19192 19200
> 0.60  19008 19200  19102 19200
> 0.80  17061 19199  18541 19198
> 0.85  14466 19193  17872 19196
> 0.90  10491 19137  15786 19181

===== VALID SPOTS* =====
Total No. of Valid Spots -----> 14432
Percentage of Valid Spots -----> 75.2

Total No. of Genes -----> 7215
Mean No. Repetitions -----> 2 for 6596 Genes
Min. No. Repetitions -----> 1 for 579 Genes
Max. No. Repetitions -----> 28 for 1 Genes

          Log(R/G) vs 0.5*Log(R*G)
          N            14432            14432
Mean      -0.017         10.327
Std        0.617         2.079
Min       -8.711         3.246
Max        4.030        15.994
Correlation 0.362

          Log(R/G) across Intensity Values
          Intensity Spots % <0 % >0
-----
( 0 , 4)           4 100.0  0.0
( 4 , 8)          1499  74.1 25.9
( 8 , 12)         9846  40.4 59.6
(12 , 16)         3083  17.3 82.7
```

Valid Spots:
 - Foreground > Background
 - Quality Flag = 0

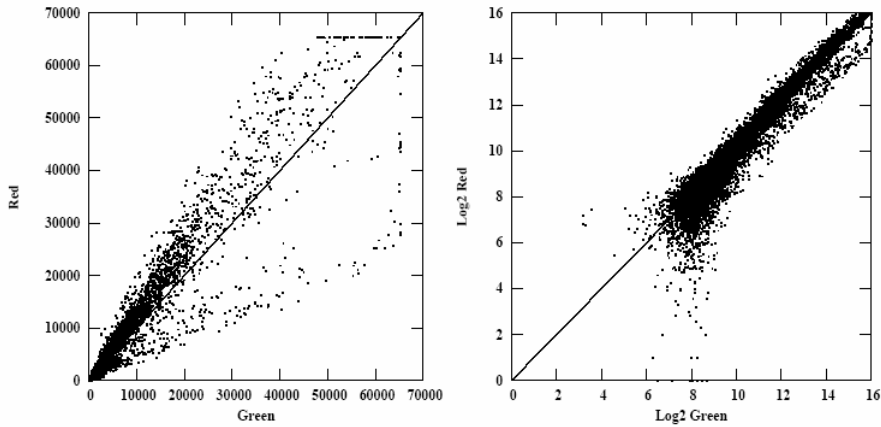
===== VALID SPOTS* =====



Diagnostic Plots

NB: Only for Valid Spot:
- Foreground > Background
- Quality Flag = 0

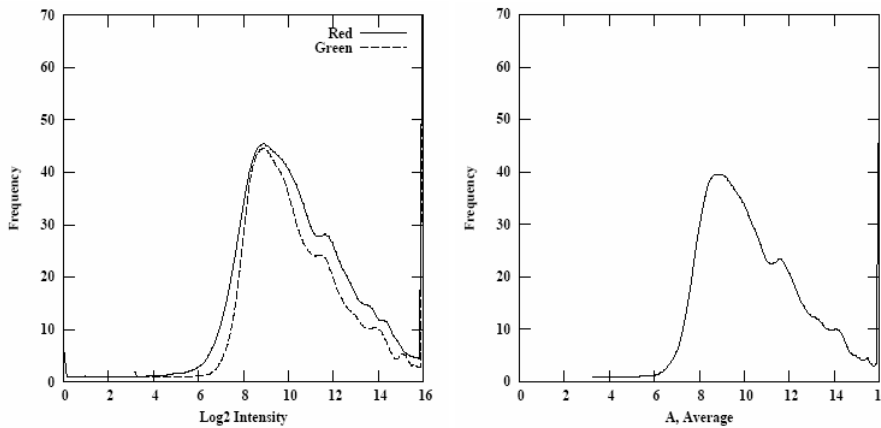
Scatters: **RED** versus **GREEN**



Diagnostic Plots

NB: Only for Valid Spot:
- Foreground > Background
- Quality Flag = 0

Densities: **RED** and **GREEN**

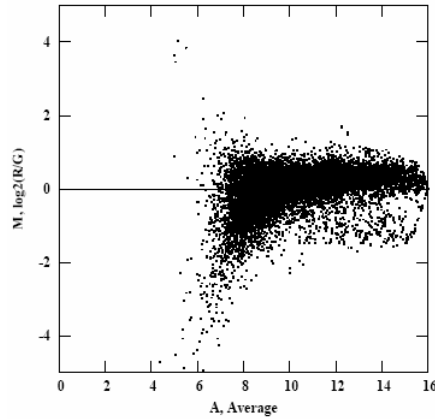
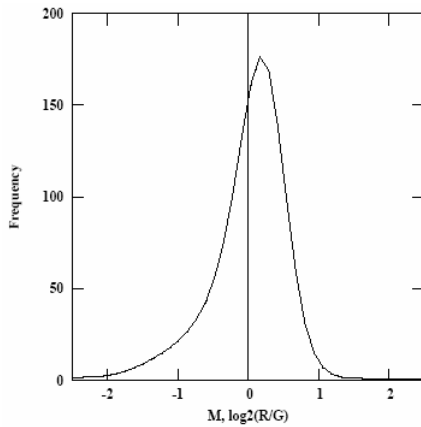




Diagnostic Plots

NB: Only for Valid Spot:
- Foreground > Background
- Quality Flag = 0

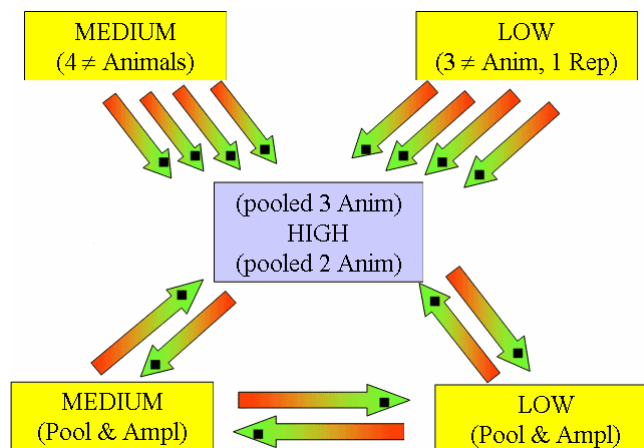
Ratios: **RED** to **GREEN**



Application

The Diets Experiment

3 Diets, 10 Animals, 14 Microarrays



Available at: <http://www.livestockgenomics.csiro.au/downloads/Keren/>