Expression Measures

Robust Multi-array Average or Robust Mult-chip Average (RMA)

The RMA method for computing an expression measure begins by computing background-corrected perfect match intensities for each perfect match cell on every GeneChip. The background corrected intensities are computed in such a way that all background-corrected values must be positive. The exact details of the background-correction method are unpublished; there is a somewhat vague description in Irizarry et al. (2003).

After background correction, the base-2 logarithm of each background-corrected perfect-match intensity is obtained. These background-corrected and log-transformed perfect-match intensities are normalized using the quantile normalization method developed by Bolstad et al. (2003). In the quantile normalization method, the highest background-corrected and log-transformed perfect-match intensity on each GeneChip is determined. These values are averaged, and the individual values are replaced by the average. This process is repeated with what were originally the second highest background-corrected and log-transformed perfect-match intensities on each GeneChip, the third highest, etc. A simplified example is given below for three hypothetical GeneChips with two probe sets and three perfect-match probes per probe set.

		Background-Corrected and Log-Transformed Perfect-Match Intensity		
Probe Set	Probe	GeneChip 1	GeneChip 2	GeneChip3
1	1	7	9	19
1	2	3	5	14
1	3	2	6	11
2	1	4	8	8
2	2	10	11	16
2	3	12	10	15

Replace the highest value on each chip with the average of the highest values.
(12+11+19)/3=14

		Background-Corrected and Log-Transformed Perfect-Match Intensity		
Probe Set	Probe	GeneChip 1	GeneChip 2	GeneChip3
1	1	7	9	**14**
1	2	3	5	14
1	3	2	6	11
2	1	4	8	8
2	2	10	**14**	16
2	3	**14**	10	15

Replace the second highest value on each chip with the average of the second highest values.
(10+10+16)/3=12

		Background-Corrected and Log-Transformed Perfect-Match Intensity		
Probe Set	Probe	GeneChip 1	GeneChip 2	GeneChip3
1	1	7	9	*14*
1	2	3	5	14
1	3	2	6	11
2	1	4	8	8
2	2	**12**	*14*	**12**
2	3	*14*	**12**	15

Replace the third highest value on each chip with the average of the third highest values.
(7+9+15)/3=10.33333

		Background-Corrected and Log-Transformed Perfect-Match Intensity		
Probe Set	Probe	GeneChip 1	GeneChip 2	GeneChip3
1	1	**10.33333**	**10.33333**	*14*
1	2	3	5	14
1	3	2	6	11
2	1	4	8	8
2	2	*12*	*14*	*12*
2	3	*14*	*12*	**10.33333**

Replace the fourth highest value on each chip with the average of the fourth highest values.
(4+8+14)/3=8.66667

		Background-Corrected and Log-Transformed Perfect-Match Intensity		
Probe Set	Probe	GeneChip 1	GeneChip 2	GeneChip3
1	1	*10.33333*	*10.33333*	*14*
1	2	3	5	**8.66667**
1	3	2	6	11
2	1	**8.66667**	**8.66667**	8
2	2	*12*	*14*	*12*
2	3	*14*	*12*	*10.33333*

Replace the fifth highest value on each chip with the average of the fifth highest values.
(3+6+11)/3=6.66667

		Background-Corrected and Log-Transformed Perfect-Match Intensity		
Probe Set	Probe	GeneChip 1	GeneChip 2	GeneChip3
1	1	*10.33333*	*10.33333*	*14*
1	2	**6.66667**	5	*8.66667*
1	3	*2*	**6.66667**	**6.66667**
2	1	*8.66667*	*8.66667*	8
2	2	*12*	*14*	*12*
2	3	*14*	*12*	*10.33333*

Replace the sixth highest value on each chip with the average of the sixth highest values. (2+5+8)/3=5

		Background-Corrected and Log-Transformed Perfect-Match Intensity		
Probe Set	Probe	GeneChip 1	GeneChip 2	GeneChip3
1	1	*10.33333*	*10.33333*	*14*
1	2	*6.66667*	**5**	*8.66667*
1	3	**5**	*6.66667*	*6.66667*
2	1	*8.66667*	*8.66667*	**5**
2	2	*12*	*14*	*12*
2	3	*14*	*12*	*10.33333*

Following quantile normalization, an additive linear model is fit to the normalized data to obtain an expression measure for each probe on each GeneChip. The linear model for a particular probe set can be written as

$$Y_{ij}=m_i+a_j+e_{ij}$$

where Y_{ij} denotes the normalized probe value corresponding to the i^{th} GeneChip and the j^{th} probe within the probe set, m_i denotes the log-scale expression for the probe set in the sample hybridized to the i^{th} GeneChip, a_j denotes the probe affinity effect for the j^{th} probe within the probe set, and e_{ij} denotes a random error term. Tukey's median polish is used to obtain estimates of the m_i values. These estimates serve as the log-scale expression measures associated with the particular probe set. We now illustrate how Tukey's median polish could be used to obtain estimates of expression for some hypothetical normalized probe set data.

		Probe					
GeneChip		**1**	**2**	**3**	**4**	**5**	**Row Medians**
	1	18	11	8	21	4	11
	2	13	7	5	16	7	7
	3	15	6	7	16	6	7
	4	19	15	12	18	5	15

Subtract the row median from each value.

		Probe				
GeneChip		**1**	**2**	**3**	**4**	**5**
	1	7	0	-3	10	-7
	2	6	0	-2	9	0
	3	8	-1	0	9	-1
	4	4	0	-3	3	-10
Column Medians		6.5	0	-2.5	9	-4

Subtract the column median from each value.

		Probe				
GeneChip		**1**	**2**	**3**	**4**	**5**	**Row Medians**
	1	0.5	0	-0.5	1	-3	0
	2	-0.5	0	0.5	0	4	0
	3	1.5	-1	2.5	0	3	1.5
	4	-2.5	0	-0.5	-6	-6	-2.5

Subtract the row median from each value.

		Probe				
GeneChip		**1**	**2**	**3**	**4**	**5**
	1	0.5	0	-0.5	1	-3
	2	-0.5	0	0.5	0	4
	3	0	-2.5	1	-1.5	1.5
	4	0	2.5	2	-3.5	-3.5
Column Medians		0	0	0.75	-0.75	-0.75

Subtract the column median from each value.

		Probe					
GeneChip		**1**	**2**	**3**	**4**	**5**	**Row Medians**
	1	0.5	0	-1.25	1.75	-2.25	0
	2	-0.5	0	-0.25	0.75	4.75	0
	3	0	-2.5	0.25	-0.75	2.25	0
	4	0	2.5	1.25	-2.75	-2.75	0

Note that row medians are all zero (above) and column medians are all zero (below).

		Probe				
GeneChip		**1**	**2**	**3**	**4**	**5**
	1	0.5	0	-1.25	1.75	-2.25
	2	-0.5	0	-0.25	0.75	4.75
	3	0	-2.5	0.25	-0.75	2.25
	4	0	2.5	1.25	-2.75	-2.75
Column Medians		0	0	0	0	0

Subtract the final values from the original values to obtain fitted values.

Fitted Values		Probe				
GeneChip		**1**	**2**	**3**	**4**	**5**
	1	17.5	11	9.25	19.25	6.25
	2	13.5	7	5.25	15.25	2.25

3	15	8.5	6.75	16.75	3.75
4	19	12.5	10.75	20.75	7.75

Now it is straightforward to see that each value in the table can be obtained by summing an estimated probe affinity with an estimated GeneChip-specific log-scale expression value. The estimated probe affinities are 4.85, -1.65, -3.4, 6.6, and -6.4 for probes 1 through 5, respectively. The estimated GeneChip-specific log-scale expression values are 12.65, 8.65, 10.15, and 14.15 for GeneChips 1 through 4, respectively. These estimated GeneChip-specific log-scale expression values would be reported as the RMA measures of expression for this hypothetical probe set.

MAS 5.0 Signal

To be described later.

References

Bolstad, B.M., Irizarry R. A., Astrand, M., and Speed, T.P. (2003). A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Bias and Variance. *Bioinformatics* 19(2):185-193.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., Speed, T. P. (2003). Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. Accepted for publication in *Biostatistics*.

Tukey, J. W. (1977). *Exploratory Data Analysis.* Addison-Wesley, Reading, Massachusetts.