Tanesha Donaldson

February 9, 2022

Bioinformatics

Assignment 2

Techniques of Normalization

When one alter/change microarray data for effects that appear from variation in the technology instead of from biological differences between the RNA samples or between the printed probes that affect the measured gene expression levels it is known as Normalization (Smyth and Speed 2003). In order to account for technical variation between the arrays, microarray needs to be normalized. Within normalization techniques, there are probes known as Mis-Match (MM) probes and Perfect match (PM) probes. PM probe sets have oligonucleotides attached to the array surface that bind the fluorescently labeled sample if transcripts corresponding to particular probe sets are expressed. However, MM oligos are made with a planned mismatch and are used to define background signal (Robson 2008). There are several techniques of normalization but in this document, I will only be discussing two.

The first technique of normalization is RMA normalization. RMA normalization is known as Robust Multichip/Multiarray Average normalization. Without the use of the information obtained in the MM probes for microarrays, RMA normalization background corrects, normalizes and summarizes the probe level information extracted from .CEL files. RMA does not use MM probes because their intensities are often higher than the match probes, making them unreliable as indicators of non-specific binding. RMA however, includes steps for background correction, quantile normalization across arrays, a probe-level model fit to each probe set across multiple arrays, and quality assessment (Pevsner 2015). The RMA background correction step includes a convolution/complex model in which the observed signal for each probe set is broken into components of true signal and noise. On the other hand, the other technique known as MAS5.0 normalization, normalizes each array independently, as well as sequentially. MAS5 calculates the difference PM – MM to obtain a robust average or for the data to have the same mean, summarizing the signal from a set of probe sets that span a gene. To conclude, the two techniques have their differences. When RMA is compared to MAS5.0, it has better accuracy being that the precision is far greater. RMA uses a multichip model, but MAS5.0 normalizes each array separately but in a logical order. RMA does not use the mismatch probes, but MAS5 uses data from mismatch probes to calculate a "robust average". Lastly, RMA values are in log2 units, but MAS5 are not (Irizarry *et al.* 2003).

References:

1. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. 2003. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Research;**31**(4):e15. DOI:https://doi.org/10.1093/nar/gng015

2. Pevsner JL. 2015. Bioinformatics and functional genomics. 3rd Edition. Chichester [u.a.]:Wiley Blackwell.

3. Robson S. 2008. RMA and GC-RMA Normalisation. University of Warwick. DOI:https://warwick.ac.uk/fac/sci/moac/people/students/2003/sam\_robson/usergroups/rmavsmas5/

4. Smyth GK, Speed T. 2003. Normalization of cDNA microarray data. Methods;**31**(4):265-73. DOI:10.1016/s1046-2023(03)00155-5. PMID:14597310.