Why Normalize?

In a microarray experiment, systematic biases such as variations in replicate slides, hybridization conditions, scanning conditions, and dye incorporation can cause global variation. Normalization removes or minimizes this variation in signal intensity levels, so that true differences in gene expression can be detected. There are various methods for normalizing data A few examples are MAS 5.0, RMA, and dChip. Each method takes a different approach at removing biases.

Many chips include perfect match probes and mismatch probes to control for cross hybridization. MAS 5.0 compares perfect match and mismatch probes and adjusts the mismatch probe value, so that it is never bigger then the perfect match probe value. MAS 5.0 divides the microarray into 16 regions. In each region, the intensity of the dimmest 2% of features is used to define the background level. Each probe is then adjusted by a weighted average of these 16 values.

RMA takes a different approach to background correction. First, only perfect match values are adjusted, the mismatched values are not changed. This is because there are too many cases where mismatch values are higher than perfect match values, which introduces more variability then the correction is worth. Second, they try to model the distribution of perfect match intensities as a sum of exponential signal with mean λ and normal noise with mean µ and variance σ 2

dChip normalizes based on probes that do not change between data such as housekeeping genes and control probes. It computationally ranks invariant probes and fits the median curve to the invariant ranked probes to adjust all value accordingly. It optionally uses either both perfect match and mismatch values or perfect match-only. Focusing on the perfect match − mismatch differences, this model estimates the probe affinities and the expression values.