

## **Data Interpretation and Choosing Candidates Based on the Analyses**

All the analyses lead to select candidate organisms that are present in some tumor samples in high number (outlier analysis of accession analysis, specific and conserved probe analysis).

Or, are present among most tumor samples (t-test of accession analysis, specific and conserved probe analysis, also MAT analysis).

Prioritize the organism candidates: method used weighted score and percent detection.

To calculate weighted score for each probe in each sample the following is done:

- the normalized g-r value for each probe in each sample if more than 3000 is multiplied by 10, if between 300 to 3000 is multiplied by 5 and if between 30 to 300 is multiplied by 1.

Next the weighted score for each probe in each sample is then summed up for a particular candidate organism. This is Weighted score sum for the particular candidate accession.

Also, calculate the prevalence of the accession across all the samples tested.

Sort candidates by percent prevalence and by the total weighted score sum per accession

Chose best probes for the top priority organisms.

Thus, it is important to look for those probes that are present in most tumor samples and are present in high number and has high probe-target binding signal.

Probes of certain organisms that are seen to be associated only with the tumor and not with the control samples will be selected.

Those probes will be used to capture the hybridized DNA from samples, which will then be sequenced in order to confirm the association of the organism to tumor.