

Supplementary data to manuscript CAN-06-0199

Identification of Specific Gene Copy Number Changes in Asbestos-Related Lung Cancer

Nymark P, et al.

Supplementary description of data processing (manuscript page 6)

Data processing: Our image analysis for detection of faulty measurement spots was performed as described previously (1) except that the spot foreground and background areas were obtained as a result of fitting a mixture of two Gaussian distributions to each spot pixel neighborhood by using an expectation-maximization (EM)-algorithm (2). In this study, the quality assessment criteria for spots included in the subsequent analysis were as follows: 1) the size of the spot is larger than 15 pixels, 2) the intensity difference of the medians of the foreground and background pixels is at least 50 and 3) the median value of local background is less than 170. These quality assessment threshold values were obtained by first forming the respective distributions for good and faulty training spots labeled by an expert. The parameters were selected to minimize to probability of misclassification of the training spots (faulty spots being classified as faulty and faulty spots classified as good). After filtering, a proper signal with information of the gene locus could be obtained for 7730 to 9071 genes in the arrays. All arrays were normalized to have equal variance and mean log₂ signal ratios.

Supplementary figure legends

Supplementary figure 1.

Array CGH results of the region 5q35.3 (chr5:175,775,917-178,511,817). The mean log₂ ratios (y-axis) in 5q35.3 are shown for each sample (x-axis). In all the exposed cases (light bars) the mean

ratio shows a deletion in the region, while amplifications are more frequent in the non-exposed cases (dark bars).

Supplementary figure 2.

Array CGH results of the regions in 9q32-34. The mean log₂ ratios (y-axis) in all three 9q regions (bars from left to right in each sample represent 9q32 (chr9:112,313,201-114,085,370), 9q33.3-q34.11 (chr9:127,249,351-128,990,540) and 9q34.13-q34.3 (chr9:132,796,807-136,051,087)) are shown for each sample (x-axis). The exposed cases (light bars) show more frequent deletions than the non-exposed cases (dark bars).

Supplementary references

1. Ruosaari, S. and Hollmén, J. Image analysis for detecting faulty spots from microarray images. *Lecture Notes In Computer Science* 2002;2534: 259 - 266.
2. McLachlan, G. *The {EM} Algorithm and Extensions*: Wiley & Sons, 1996.