**Ad Statistical methods**

In a first univariate approach associations of CNA regions with dose or phenotypes or clinical parameters were tested with help of the R-package CGHtest (R statistical software, M. van de Wiel, <http://www.few.vu.nl/~mavdwiel/CGHtest.html>). In that package, p-values are adjusted by permutation based FDR-calculations. Strong overlap between the univariate results of some variables was the motivation to finally apply a multivariate approach in order to assess the statistical interaction or independent association of pairs of variables with gain or loss frequencies. The multivariate analysis was performed with logistic two-way modelling for the binary variable *y*, „gain versus non-gain“ or „loss versus non-loss“. In symbolic notation

*y ~ A + B + A:B*, (1)

where A:B represents the interaction between variables (i.e. effects) *A* and *B*. Significance for *A:B* was tested as contrast of the full model (1) to the pure additive model

*y ~ A + B*, (2)

using the deviance difference between the two models (McCullagh, Peter; Nelder, John, 1989). First, for each CNA region the significance of the full model was tested and p-values adjusted for FDR1. In a second step, the significance of the interaction term of model (1) was tested for those CNA regions, where the full model was significant. Finally, the independent additive contributions of *A* and/or *B* in model (2) were tested, if the interaction was not significant. p-values for each testing step were adjusted for FDR1 and significance was generally accepted for FDR<0.05.

Multivariate analysis was performed for the three possible pairs A, B of variables radiation dose, latency time and Gender. Radiation dose was treated as a categorical variable with three groups, low dose (up to 300 mGy), moderate (between 300 and 1000 mGy) and high (more than 1000 Gy). Latency time was calculated as the difference between age at operation and age at exposure and treated as binary variable with short and long latency (up to 17 years and longer than 17 years, 17 years the closest integer to the median value). Multivariate analysis with more than two variables was not performed because of too small combined group sizes.

**References:**

McCullagh, Peter; Nelder, John (1989). Generalized Linear Models, Second Edition. Boca Raton: Chapman and Hall/CRC