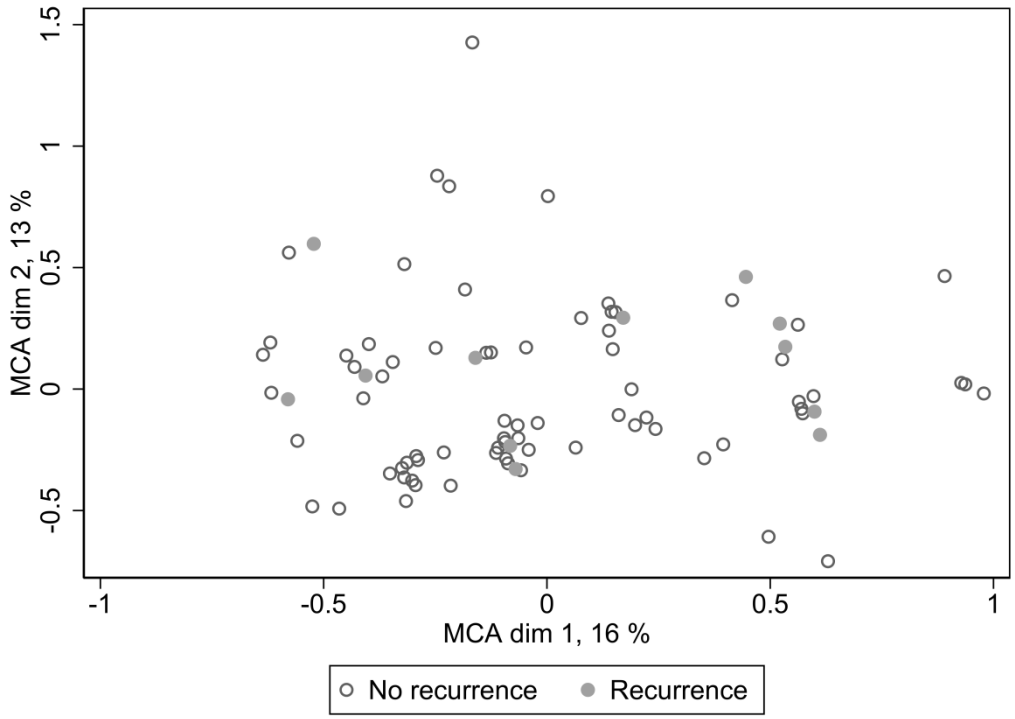

Fig. S1a.

Multiple correspondence analysis (MCA) based on the gene mutation status of each CRC-associated gene. The scatterplot shows the sample coordinates of the two first MC dimensions of the MCA, and exemplifies clustering of patients with labelling based on cancer recurrence.

The three first MC dimensions explained 16, 13 and 12% of the total variation in data, respectively. Sample coordinates of the first dimension were associated with tumor differentiation, whereas coordinates of the second dimension were associated with gender. No association with cancer recurrence or with tumor location was observed. *BRAF*, *APC* and *KRAS* contributed the most to the first dimension (cousine^2 dimension one *BRAF* = 0.62, *APC* = 0.44, *KRAS* = 0.29), whereas *PTEN*, *CTNNB1* and *MET* contributed the most to dimension two (cousine^2 dimension two *PTEN* = 0.58, *CTNNB1* = 0.41, *MET* = 0.17).

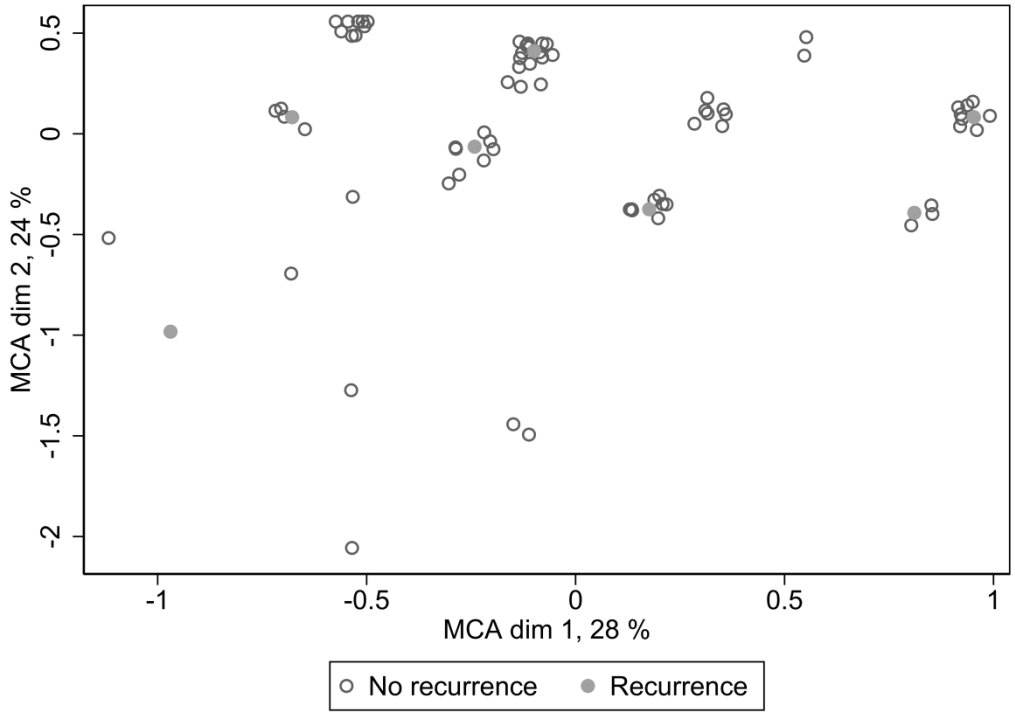

Fig. S1b.

Multiple correspondence analysis (MCA) based on the gene mutation status of the six CRC-associated genes contributing the most to dimension one and two when all CRC-genes were included; *APC*, *BRAF*, *KRAS*, *CTNNB1*, *PTEN* and *MET*. The scatterplot shows the sample coordinates of the two first MC dimensions of the MCA and exemplifies clustering of patients with labelling based on cancer recurrence. With a smaller number of genes a larger proportion of variation in data was explained by each dimension, but the statistical results remained the same as with all CRC-associated genes included in the analysis.