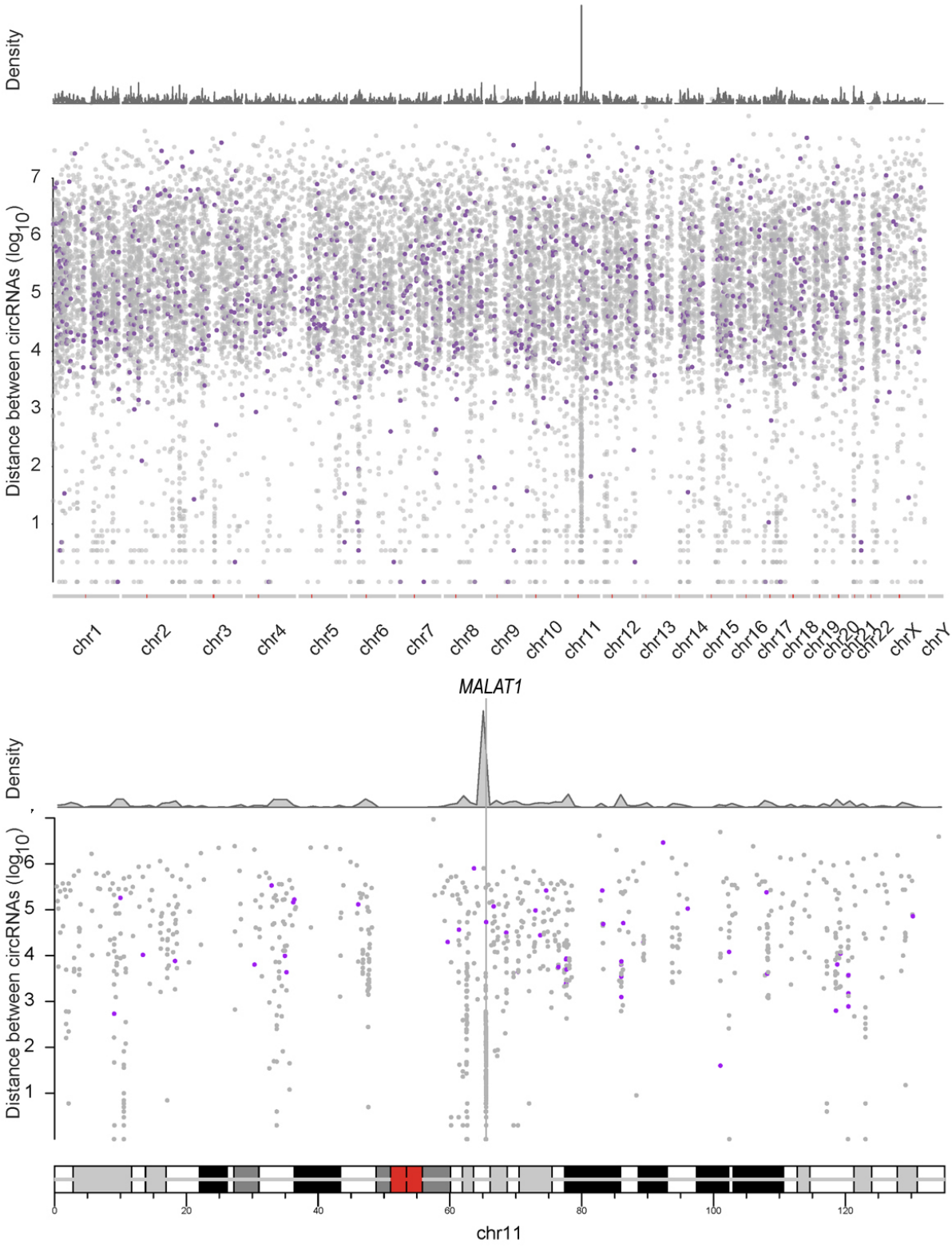
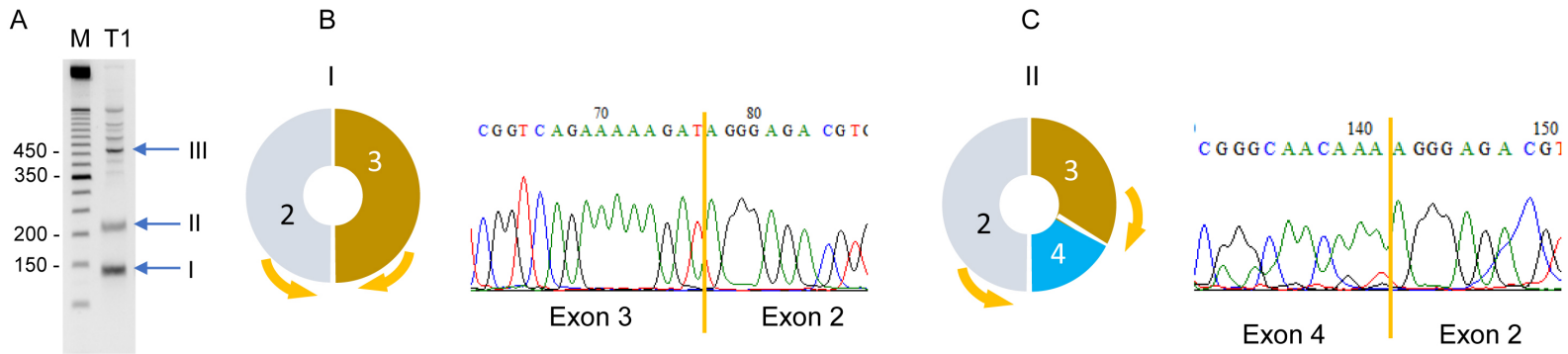
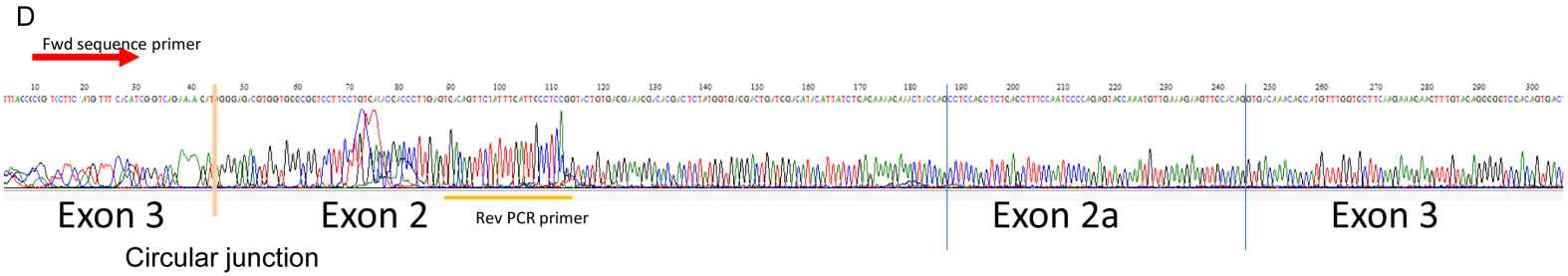
Supplemental Fig S1: Chromosomal distribution of circRNAs. In purple, rainfall-plot shows the location of the circRNAs that are identified in at least 50 samples. Top panel, genome-wide, bottom panel, Chromosome 11.

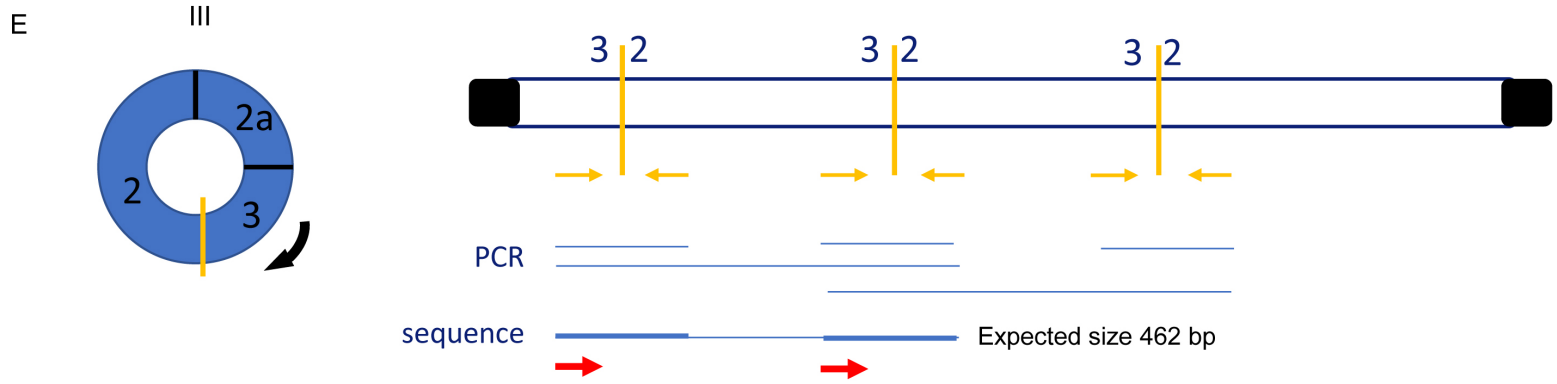


Supplemental Fig S2: Sanger sequence analysis of CNOT2

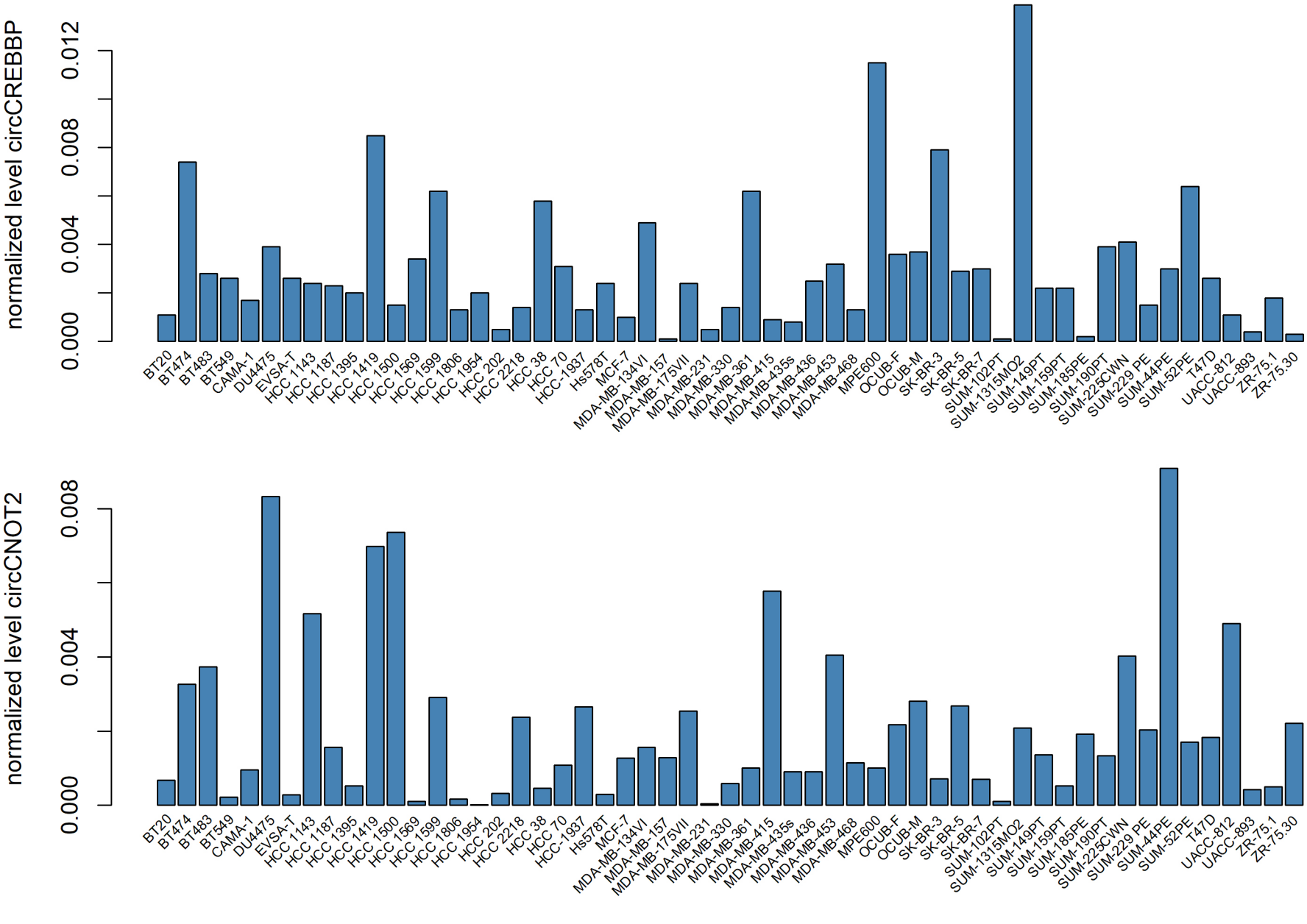
CircCNOT2 was PCR amplified, and three PCR fragments were excised from agarose-gel and Sanger sequenced. **A** Agarose gel image of PCR fragments. M indicates the marker, T1 the cDNA pool of primary breast cancers. **B,C** Schematic overview of *circCNOT2* and results of the Sanger sequencing. Gold arrows indicate the PCR primers, the vertical gold line indicates the circular junction. Expected sizes were 140 bp for fragment I and 206 for fragment II **D** Sanger sequence of the largest excised PCR fragment (III) of figure 2a. **E** Model to explain the observed sequence. During cDNA generation (left part, black arrow) the RT polymerase displaces the newly generated strand, thereby generating a linear molecule containing several copies of the circular junction (right part, vertical gold lines). PCR primers (gold arrows) around the circular junction can generate amplified fragments of several sizes. Then, during Sanger sequencing the sequence primer (red arrow) thus can yield the sequence observed in this figure 2d. The expected size is 462 bp: 142, 58 and 122 bp for exon 2, 2a and 3, respectively, plus the sizes of the primer to the circular junction: 73 and 67 bp.





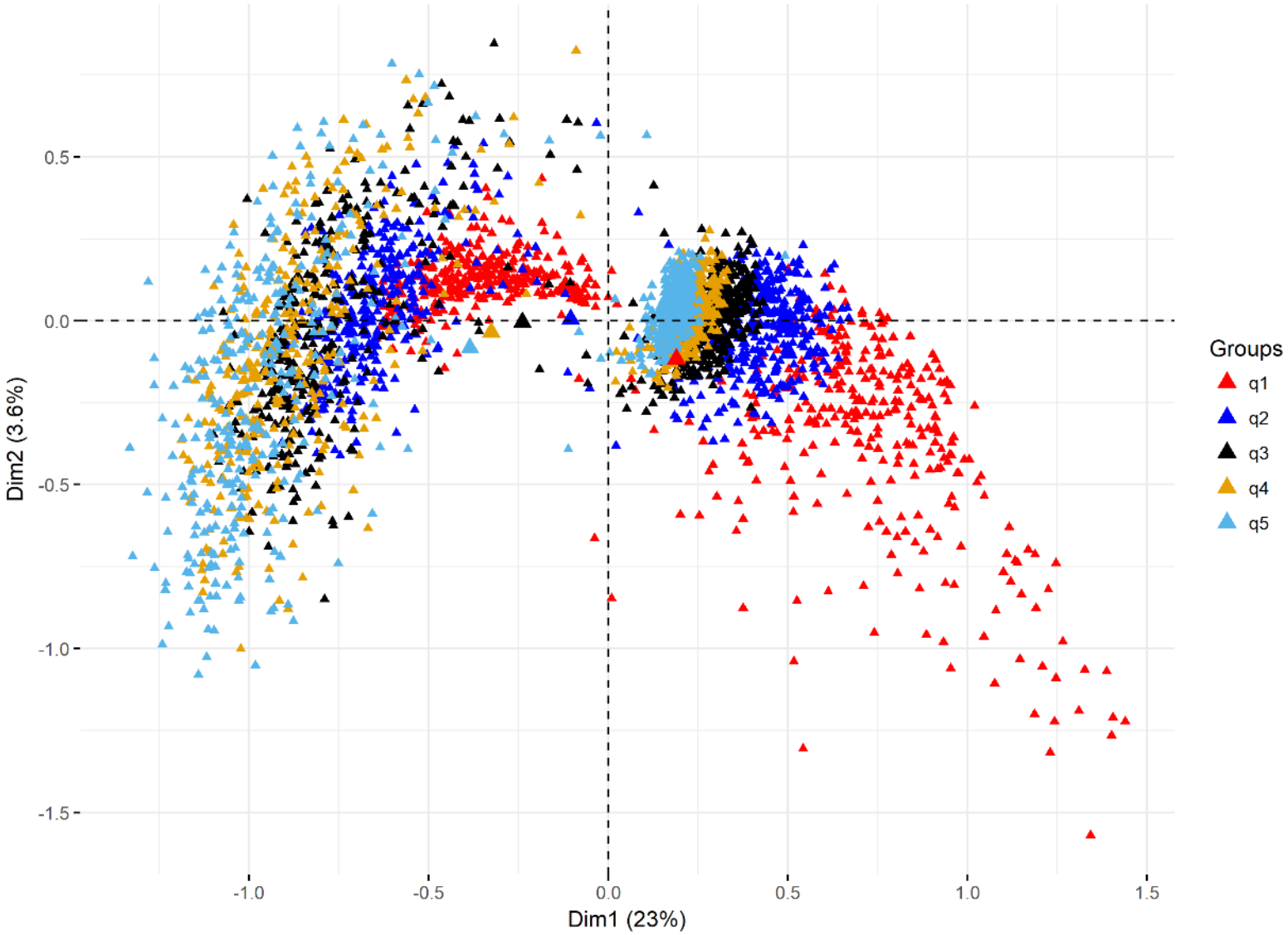


Supplemental Fig S3: Expression levels of *circCREBBP* and *circCNOT2* in 55 breast cancer cell lines.
Expression levels were quantified relative to the average expression of 3 reference genes (*HPRT1*, *HMBS* and *TBP*).



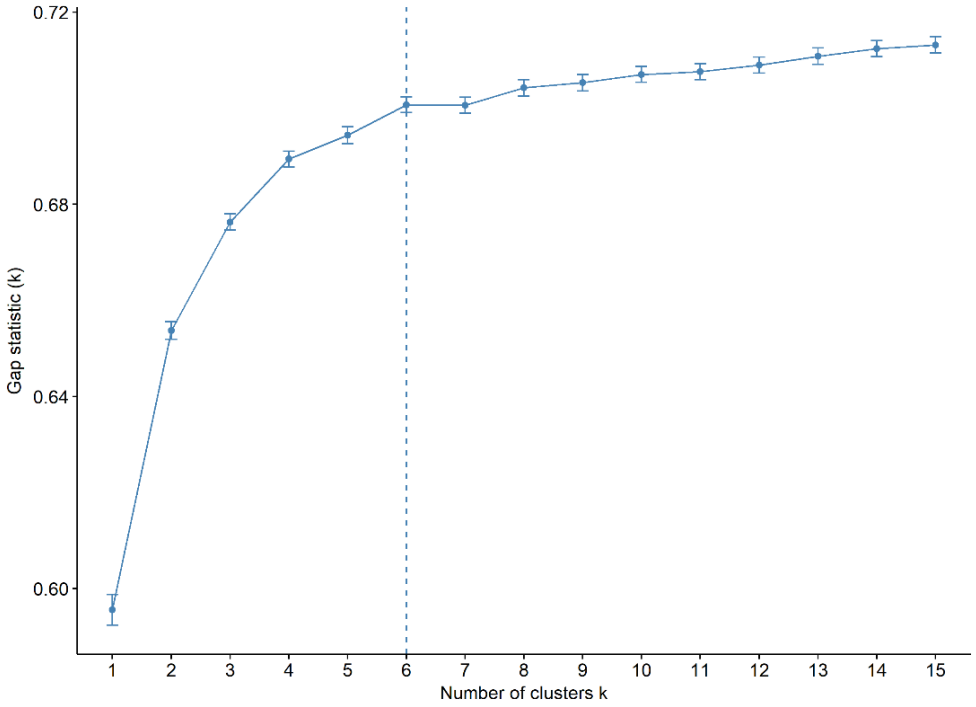
Supplemental Fig S4: MCA plot of circRNAs

The presence (mostly left of the y-axis, see figure 5 of the main article) or absence (right of the y-axis) of a circRNA of a gene is colored by recurrence – the number of times a gene has a circRNA identified in different samples. Recurrence was divided into 5 quantiles q1 to q5, with q1 the most recurrent.



Supplemental Fig S5: Optimum number of clusters

Using the Gap statistic samples were analyzed in k=1 to k=15 clusters (k-means as partitioning method).



Supplemental Fig S6: Expression level versus correlation between linear and circular RNA.

The median expression level (TMM-normalized data) of each gene (linear transcript) was calculated and plotted against the correlation coefficient between linear and circular expression.