



**Extended Data Figure 5 | Transcriptomic analysis of F3-T3 fusion-positive and fusion-like GBM and validation of ee-MWW using different cancer-driving alterations.** **a**, Consensus clustering on the Euclidean distance matrix based on the top and bottom 50 genes having the highest and lowest probability to be upregulated, respectively, in the nine F3-T3 fusion-positive GBM. The consensus matrix is obtained from 10,000 random samplings using 70% of the 544 samples. The nine F3-T3-positive GBM (in red) fall in one cluster (cyan). **b**, MWW enrichment plot of the 'hallmark oxidative phosphorylation' GO category in F3-T3-positive GBM. NES and  $P$ -values are indicated. **c**, Enrichment map network of statistically significant GO categories ( $Q < 0.001$ , NES  $> 0.6$ , upper-tailed MWW-GST) in nine fusion-like GBM. Nodes represent GO terms and lines demonstrate their connectivity. Size of nodes is proportional to number of genes in the GO category and thickness of lines indicates the fraction of genes shared between the groups. **d**, Analysis of copy number amplification of the *POLRMT* gene comparing fusion-like GBM with all other GBM at different thresholds for amplification detection on log-R

ratio from single-nucleotide polymorphism arrays.  $P$  value and log-odds at different thresholds are indicated (Fisher's exact test). **e**, *POLRMT* gene expression in fusion-like GBM ( $n = 9$ ) and the remaining samples ( $n = 535$ ). Box plot spans the first to third quartiles and whiskers show the  $1.5 \times$  interquartile range.  $P$  value, two-sided MWW test. **f**, Representative images of VDAC1 immunofluorescence in F3-T3-positive (green, left) and F3-T3-negative (right) GBM. DAPI staining of nuclei is shown as an indication of cellular density (blue, bottom). **g**, Hierarchical clustering of two GBM samples with a KRAS mutation (red) out of 544 samples. Heat map of the two KRAS mutant samples is enlarged to the left. **h**, Hierarchical clustering of five KRAS-mutated samples (red) in the invasive breast carcinoma (BRCA) cohort ( $n = 1,093$ ). **i**, Hierarchical clustering of six EGFR-SEPT14-positive GBM samples (red) out of 544 samples. Data in **g**–**i**, were obtained using the Euclidean distance and Ward linkage method and are based on the top and bottom 50 genes having the highest and lowest probability to be upregulated, respectively.