Database search (GEO, AE) of published EMT studies (replicates ≥ 2)

Gathering of processed expression data

Logarithmizing data
Calculation of FC
Two-tailed Student’s t-test

Selection of genes:
FC ≥ 2; FC ≤ 0.5
p < 0.05

Insertion of data into matrix:
Up: 1; Down: -1
Not regulated: 0

Calculating binomial distribution function
Preliminary clustering of all cutoff options

Clustering of genes shared between ≥ 10 studies

365 genes list

Selection of genes:
Up in ≥ 10
OR
Down in ≥ 10

130 genes EMT-core list

Bioinformatic single enrichment and pathway analysis

Correlation with clinical data

Figure S3: Flow chart depicting the generation of the EMT-core gene list.
Overview of the meta-analysis, the generation of the 365 genes list, the EMT-core list as well as the single enrichment analysis and the correlation with clinical data. GEO, Gene expression Omnibus; AE, Array Express; EMT, epithelial to mesenchymal transition; FC, Fold change; Up, upregulated; Down, downregulated.