Colon Cancer (TCGA-COAD) (n=338)

Marisa et al. (n=566)

CRC intrinsic subtype

FKPM_UQ

LARGE2 expression score

CRISA  CRISB  CRISC  CRISD  CRISE

CRC intrinsic subtype

CRISA  CRISB  CRISC  CRISD  CRISE

LARGE2 expression score

1 25 50

Rank-List (*1000)

Wnt-Catenin-Signaling (Hallmark, 42 genes)

Nes : 2.36
FDR-q : 0.0

Colon Cancer (TCGA-COAD)

LARGE2p.c. LARGE2n.c.

Rank-List (*1000)

Intrinsic (oncogenic) Wnt signature (112 genes ; Farin et al.)

Nes : 2.69
FDR-q : 0.0

EphB2+ hCoSC signature (59 genes ; Jung et al.)

Nes : 2.41
FDR-q : 0.0

PTK7+ hCoSC signature (327 genes ; Jung et al.)

Nes : 2.60
FDR-q : 0.0

CRC (PDXs, Isella et al.)

Nes : 2.10
FDR-q : 0.0

EPHB2+ hCoSC signature (59 genes ; Jung et al.)

Nes : 2.49
FDR-q : 0.0

PTK7+ hCoSC signature (327 genes ; Jung et al.)

Nes : 2.82
FDR-q : 0.0

CRC (PDXs, Isella et al.)

Nes : 2.10
FDR-q : 0.0

PTK7+ hCoSC signature (327 genes ; Jung et al.)

Nes : 1.81
FDR-q : 0.001

58 CRC cell lines (CCLE)
**A,B** LARGE2 expression in TCGA-COAD (A, n=338) and Marisa et al.(2013) CRC cohort (B, n=566), all stratified into CRC intrinsic subtypes (CRIS) according to Isella et al. (2017). Displayed are the total means, asterisks indicate the significance of differences between individual subtypes compared to CRISD, calculated by one-way ANOVA (**p < 0.01, *** p < 0.001, **** p < 0.0001).

**C-G** Gene set enrichment analysis (GSEA) was performed with a signature of LARGE-correlated genes derived from the TCGA-COAD and TCGA-READ data sets. Analyses used a Wnt-Catenin-Signaling gene set (Hallmark gene set, Broad Institute), an intrinsic oncogenic Wnt gene set (Michels et al., 2019) (**C-E**), and two sets of genes specifying EPHB2 positive (**F**) or PTK7 positive (**G**) human colonic stem cells (Jung et al., 2011 and 2015).

**H-J** GSEA was performed with a LARGE2 gene signature derived from the CRC Pdx dataset, published by Isella et al.(2017), using the Wnt hallmark gene set (**H**) and PTK7 positive and EPHB2 positive human colonic stem cell gene sets (**I,J**).

**K,L** GSEA was performed with a LARGE2 gene signature derived from the Cancer cell line encyclopedia (CCLE) CRC cell line (n=58) microarray dataset using the PTK7 positive and EPHB2 positive human colonic stem cell gene sets.