Supplementary Figure 7. RT-PCR versus microarray for the 30 malignancy-risk genes: prediction performance, PCA comparison, and correlation plot.

1. Prediction performance:
    a. Both microarray expression and RT-PCR expression were standardized by centered at mean and scaled by standard deviation (SD). That is, for each gene, we calculate the mean and SD among the 21 samples, then generate a standardized score, (x-mean(x))/sd(x) where x is an expression value.
    b. Classifier by support vector machine (SVM): We used the SVM to build a classifier using the 30 genes microarray expression. Leave-One-Out-Cross-Validation (LOOCV) was used to assess the quality of the classifier. The accuracy rate was 86%.
    c. Prediction: The classifier was used to evaluate PR-PCR platform. That is, the classifier converted the 30 genes RT-PCR expressions into a predicted value for each sample to determine which group each tissue belongs to. The prediction results yielded a 90% accuracy rate.

		Predicted Outcome		
		Normal	IDC-like normal	IDC
True Outcome	Normal	7	1	0
	IDC-like normal	0	4	1
	IDC	0	0	8

2. Comparison of various PCAs: We performed two ways to compare PCAs. The first one is to compare correlation of microarray and RT-PCR at each PCA. Results in Figure A showed a very high correlation in PCA1 ($r=0.95$). In contrast, the other PCAs had a weak correlation ($r<0.5$). The second approach is to test each PCA (PCA1, PCA2, and PCA3) to see if it is associated with cancer status (i.e., normal, IDC-like normal, to, IDC). Results in Figure B showed that an increasing trend of the risk score (from normal to IDC tissues) by PCA1 in the first column panel. On the other hand, the PCA2 and PCA3 did not yield their association with cancer status.

Figure A: Correlation of microarray and RT-PCR at each PCA

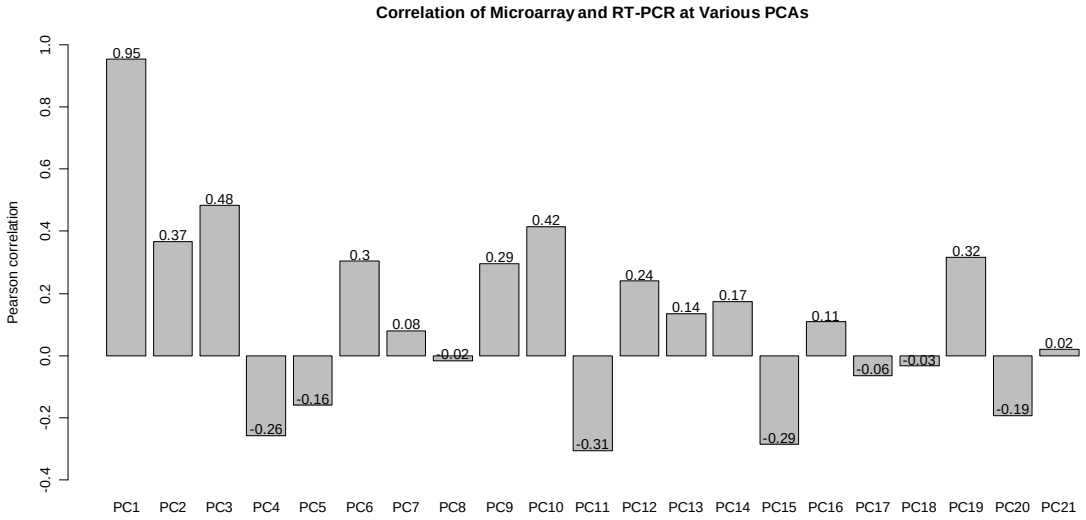
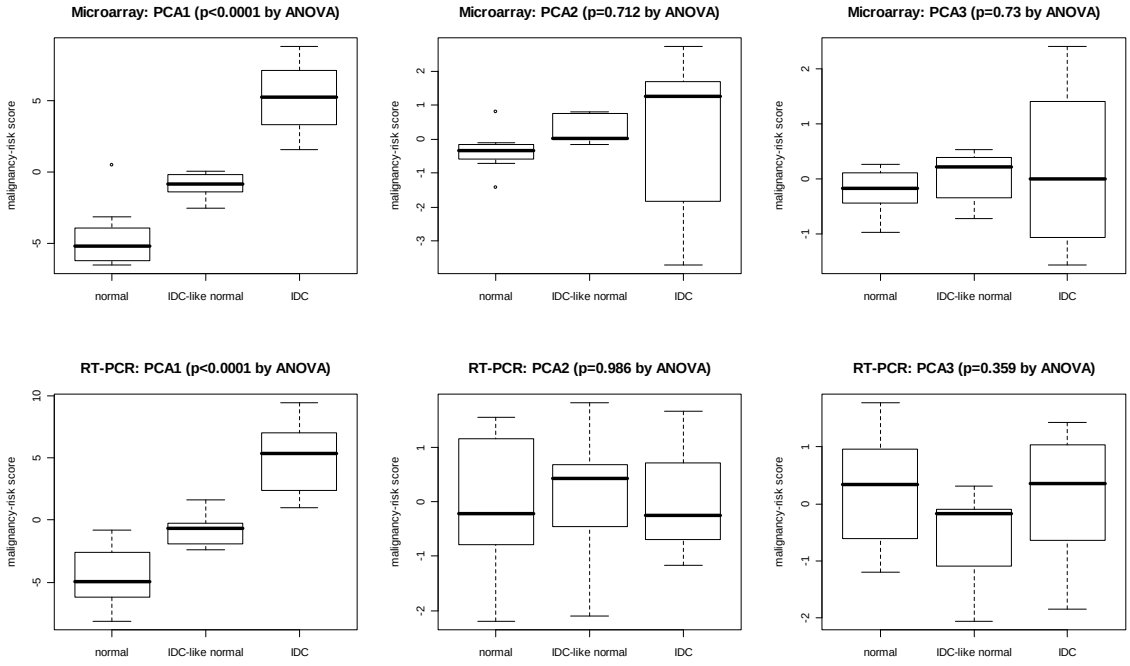
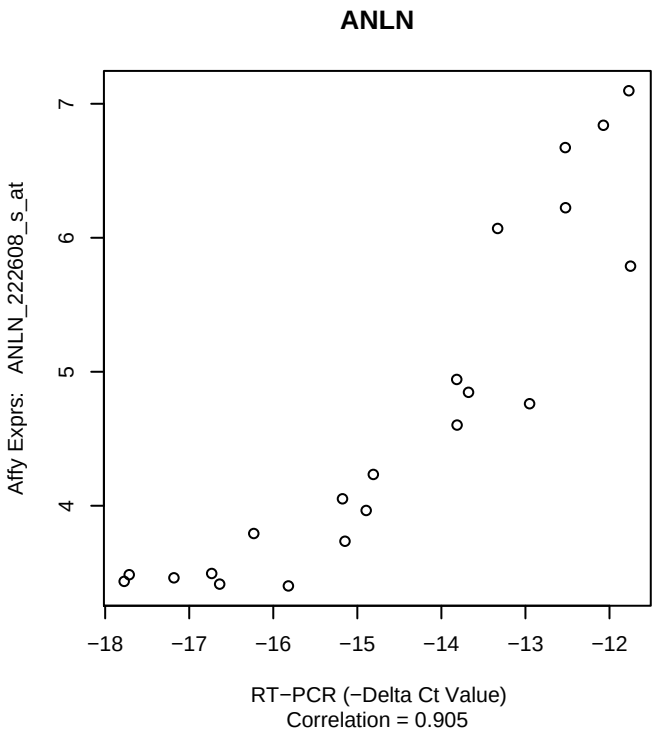
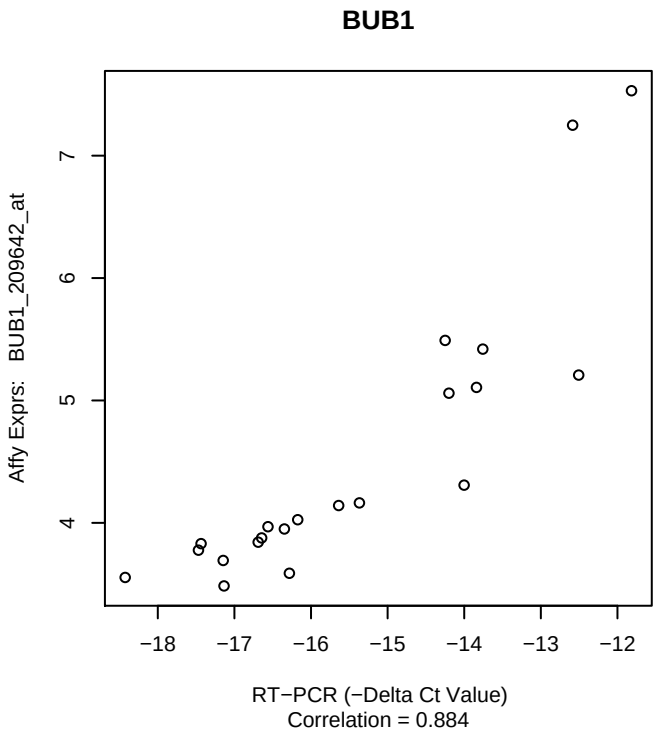


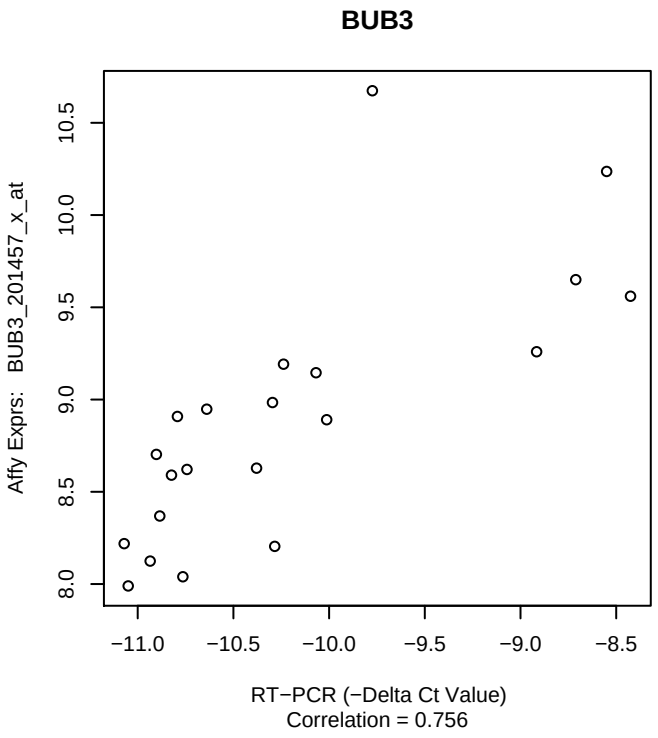
Figure B: Comparison of Various Principle Components (PCA): PCA 1 means the malignancy-risk score was generated by the 1st PCA. We did the same way for PCA2 and PCA3 in microarray and RT-PCR, respectively. The PCA1 (microarray and RT-PCR) demonstrated a trend of the risk score from normal to IDC tissues in the first column panel. In contrast, the PCA2 and PCA3 did not show its association with cancer status.

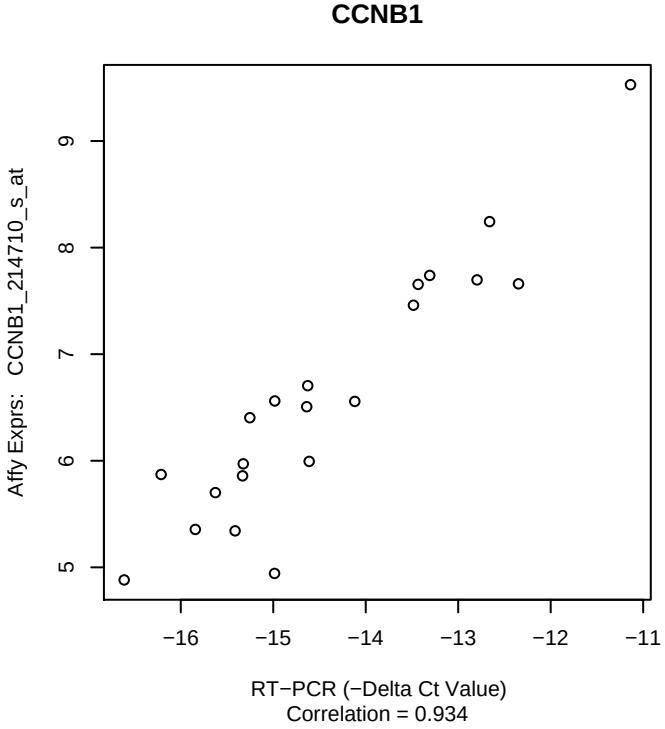


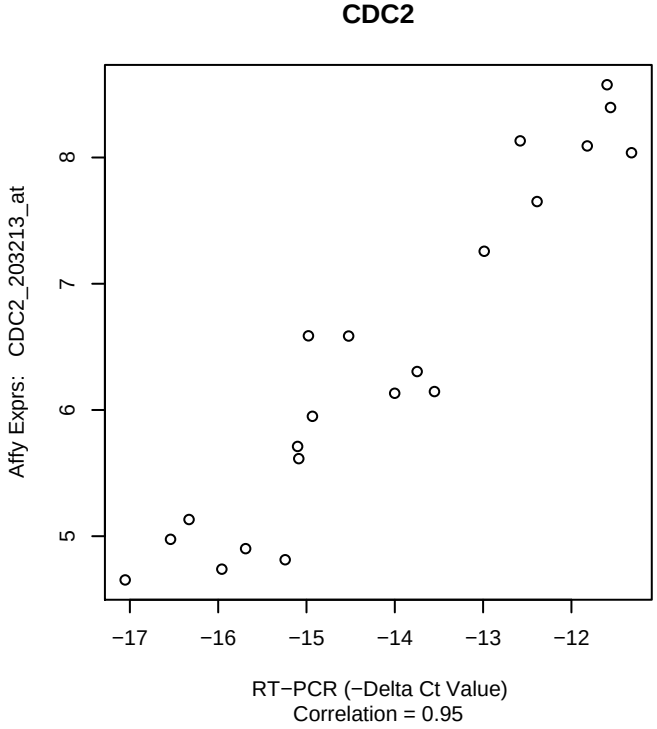
3. Correlation plot:

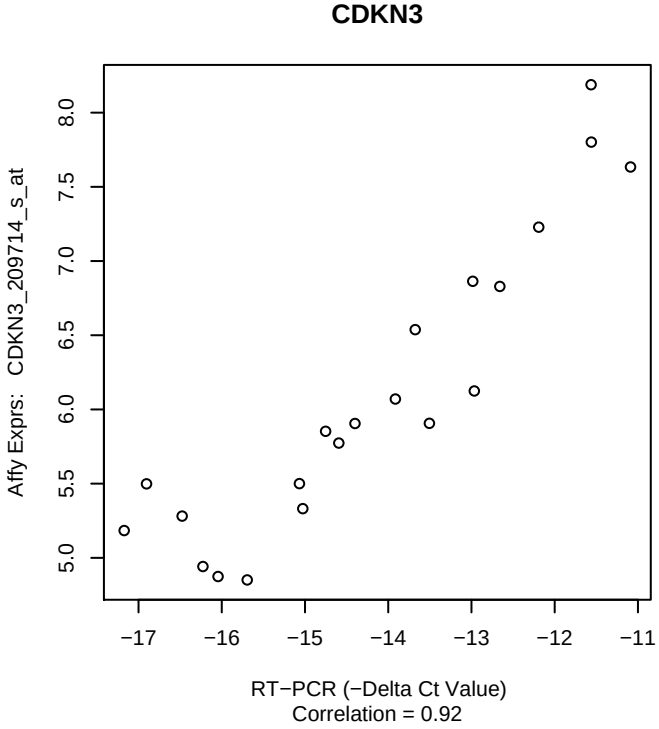

ANLN
Affy Exprs: ANLN_222608_s_at
RT−PCR (−Delta Ct Value)
Correlation = 0.905

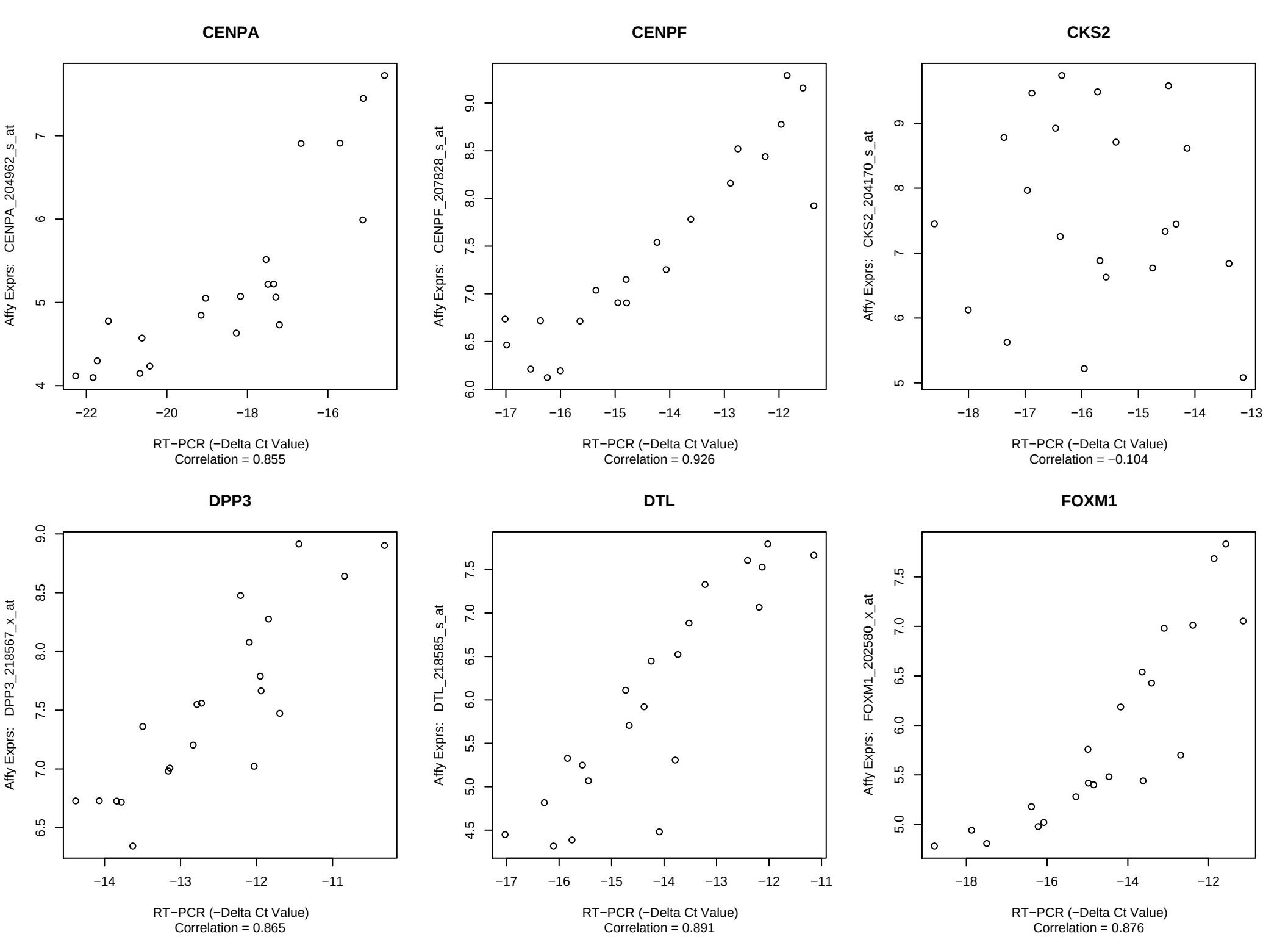

BUB1
Affy Exprs: BUB1_209642_at
RT−PCR (−Delta Ct Value)
Correlation = 0.884

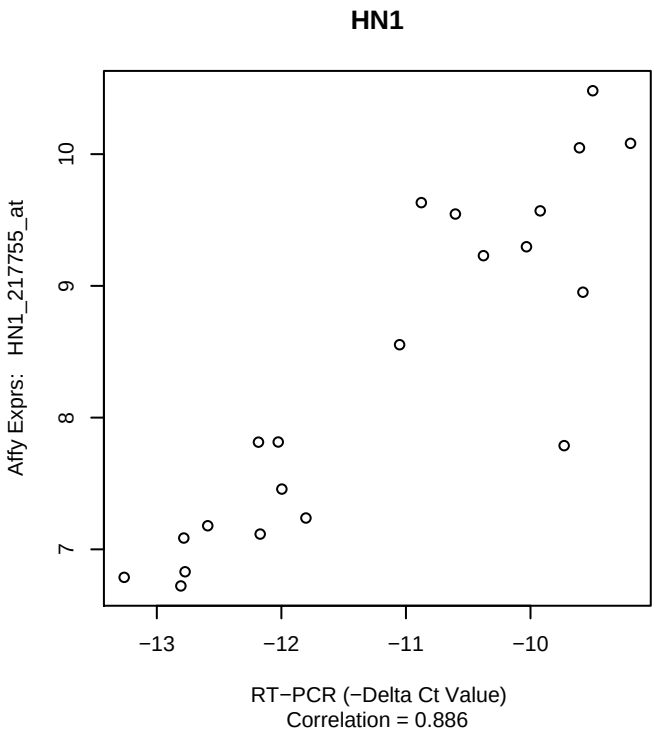

BUB3
Affy Exprs: BUB3_201457_x_at
RT−PCR (−Delta Ct Value)
Correlation = 0.756

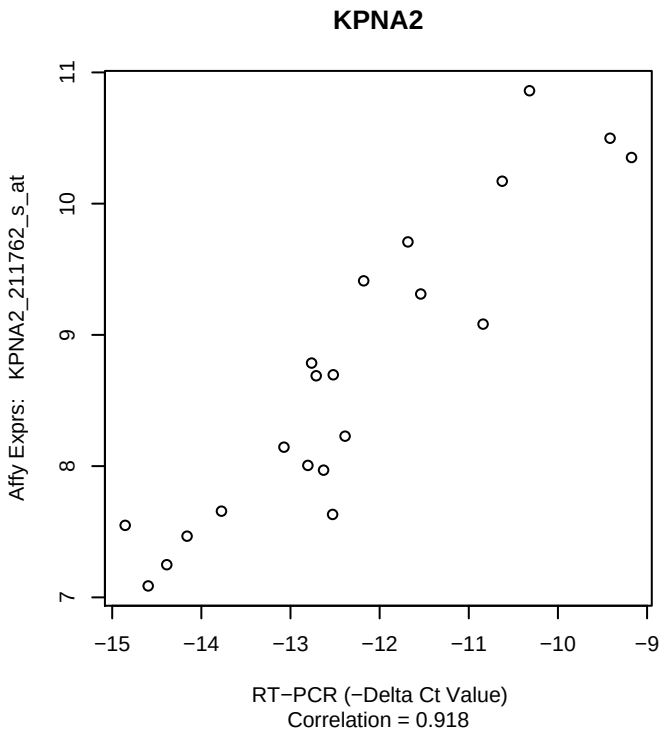

CCNB1
Affy Exprs: CCNB1_214710_s_at
RT−PCR (−Delta Ct Value)
Correlation = 0.934

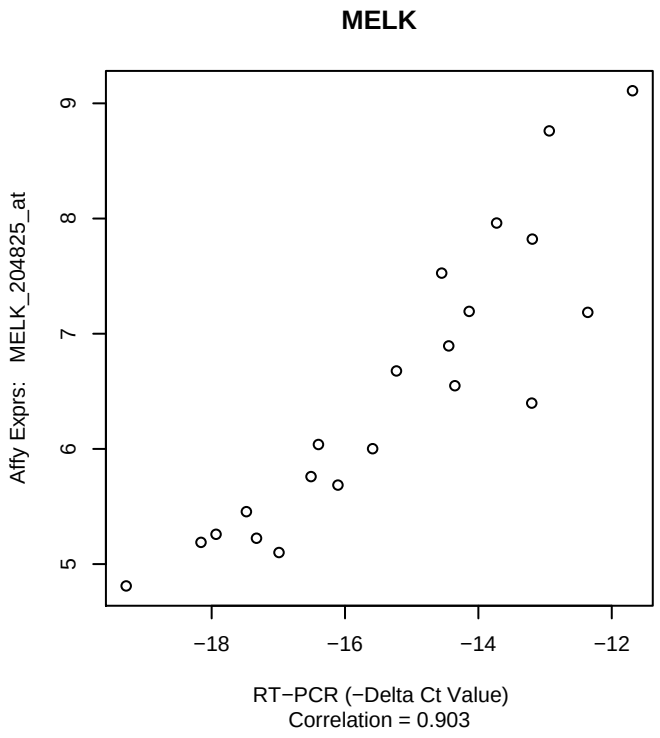

CDC2
Affy Exprs: CDC2_203213_at
RT−PCR (−Delta Ct Value)
Correlation = 0.95

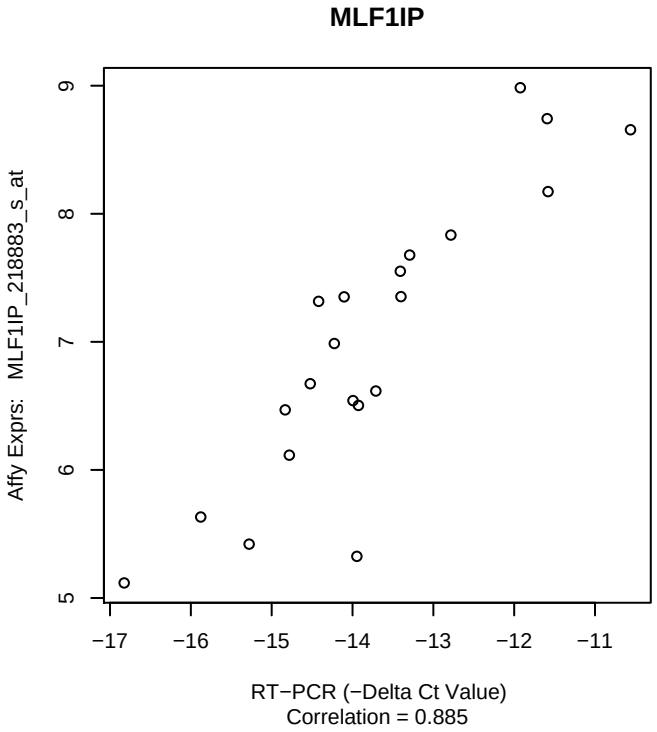

CDKN3
Affy Exprs: CDKN3_209714_s_at
RT−PCR (−Delta Ct Value)
Correlation = 0.92

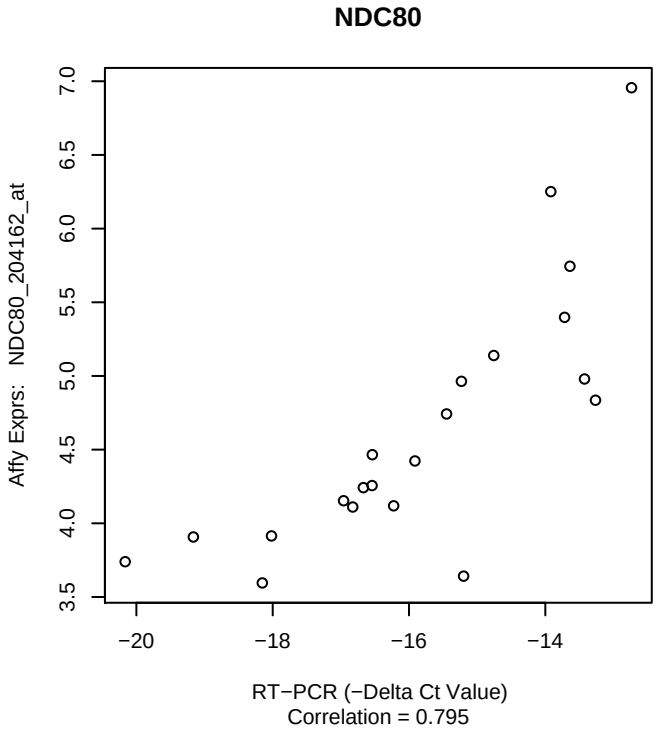

CENPA
Affy Exprs: CENPA_204962_s_at
RT−PCR (−Delta Ct Value)
Correlation = 0.855
CENPF
Affy Exprs: CENPF_207828_s_at
RT−PCR (−Delta Ct Value)
Correlation = 0.926
CKS2
Affy Exprs: CKS2_204170_s_at
RT−PCR (−Delta Ct Value)
Correlation = −0.104
DPP3
Affy Exprs: DPP3_218567_x_at
RT−PCR (−Delta Ct Value)
Correlation = 0.865
DTL
Affy Exprs: DTL_218585_s_at
RT−PCR (−Delta Ct Value)
Correlation = 0.891
FOXM1
Affy Exprs: FOXM1_202580_x_at
RT−PCR (−Delta Ct Value)
Correlation = 0.876

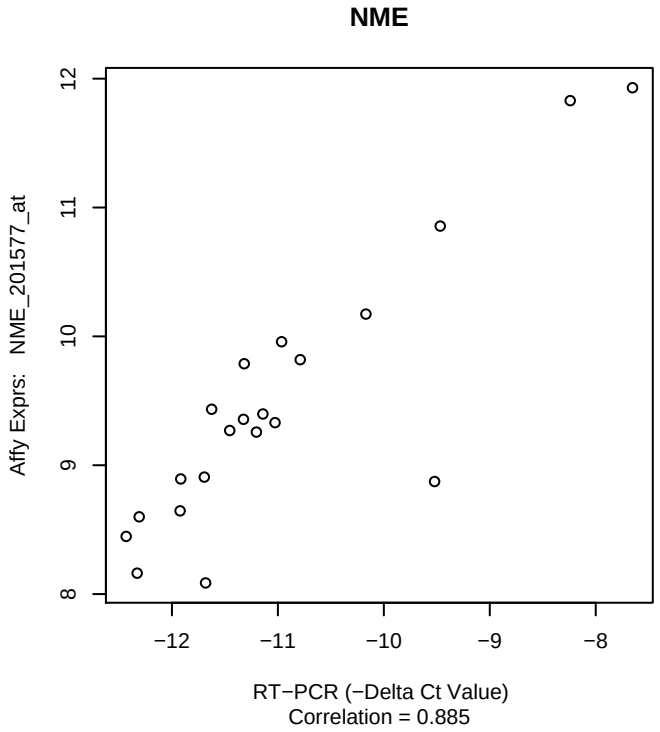

HN1
Affy Exprs: HN1_217755_at
RT−PCR (−Delta Ct Value)
Correlation = 0.886

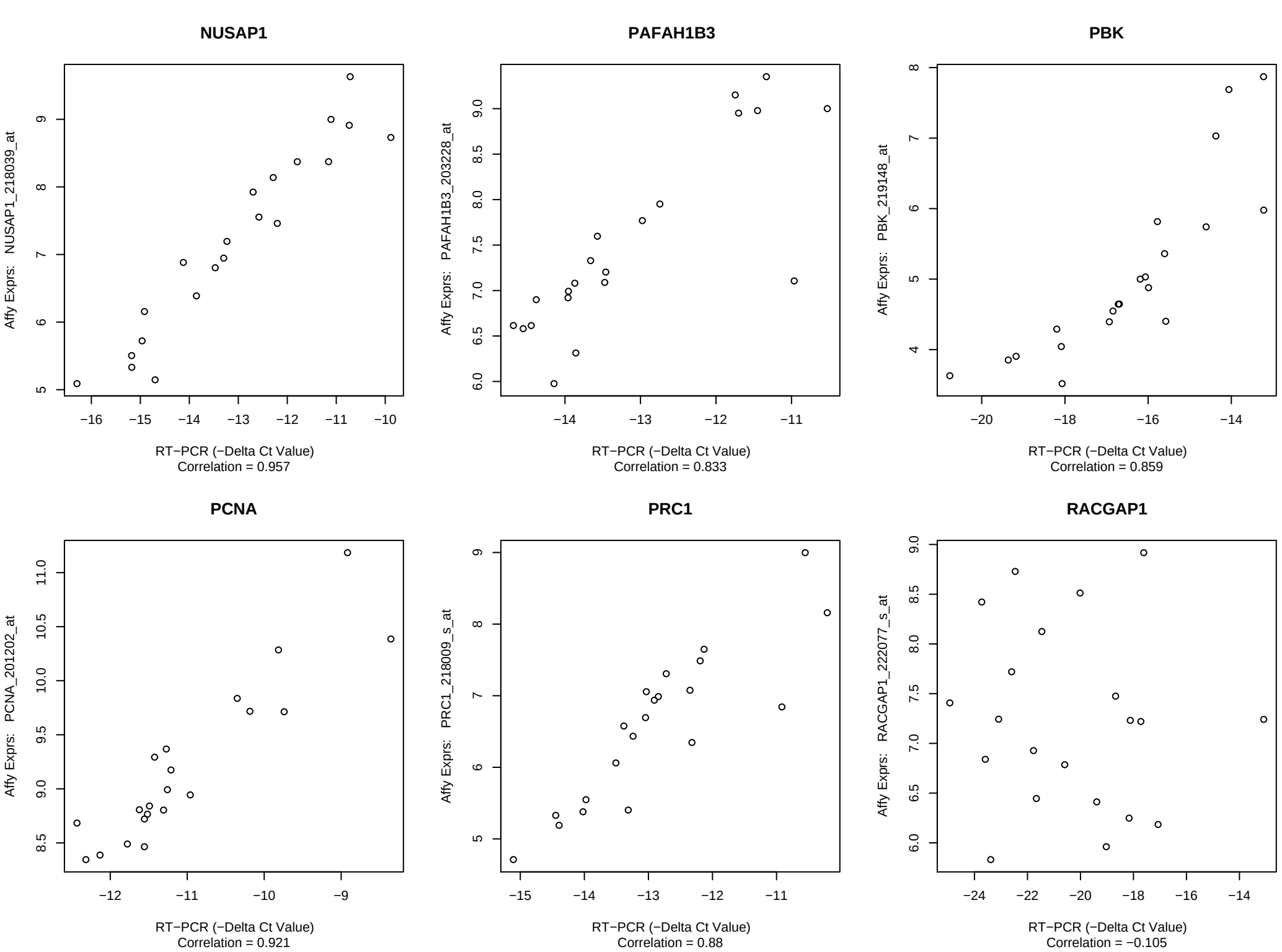

KPNA2
Affy Exprs: KPNA2_211762_s_at
RT−PCR (−Delta Ct Value)
Correlation = 0.918

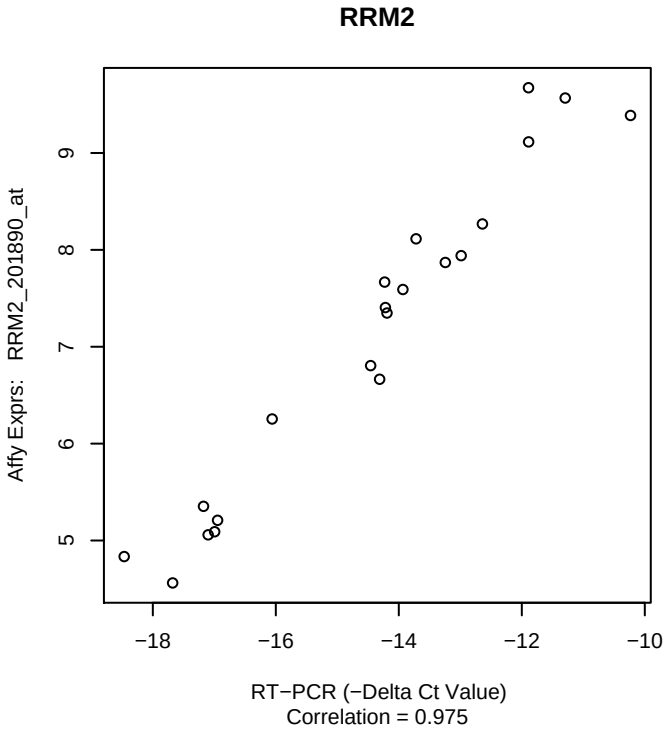

MELK
Affy Exprs: MELK_204825_at
RT−PCR (−Delta Ct Value)
Correlation = 0.903

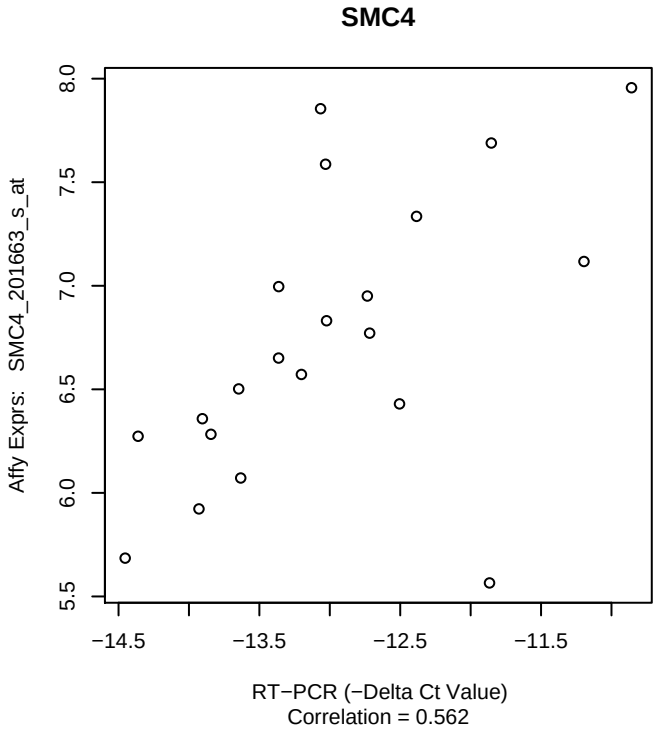

MLF1IP
Affy Exprs: MLF1IP_218883_s_at
RT−PCR (−Delta Ct Value)
Correlation = 0.885

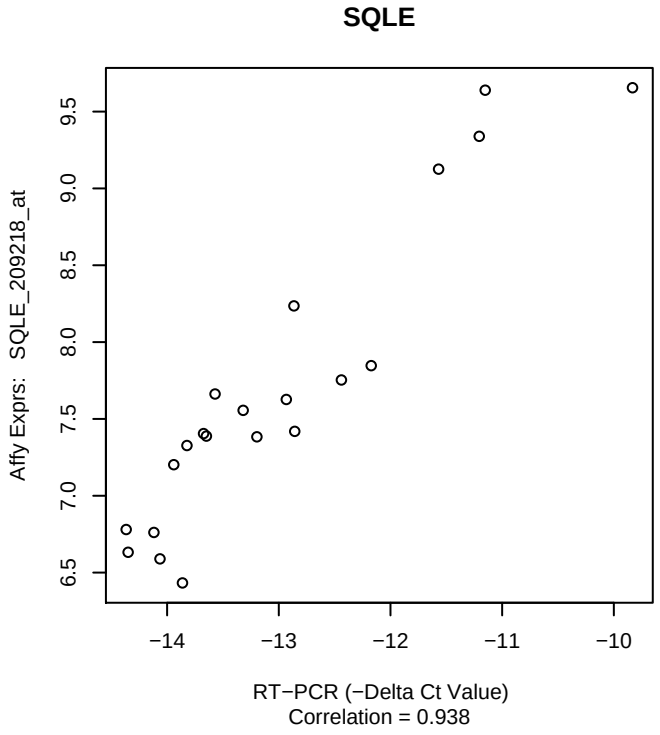

NDC80
Affy Exprs: NDC80_204162_at
RT−PCR (−Delta Ct Value)
Correlation = 0.795

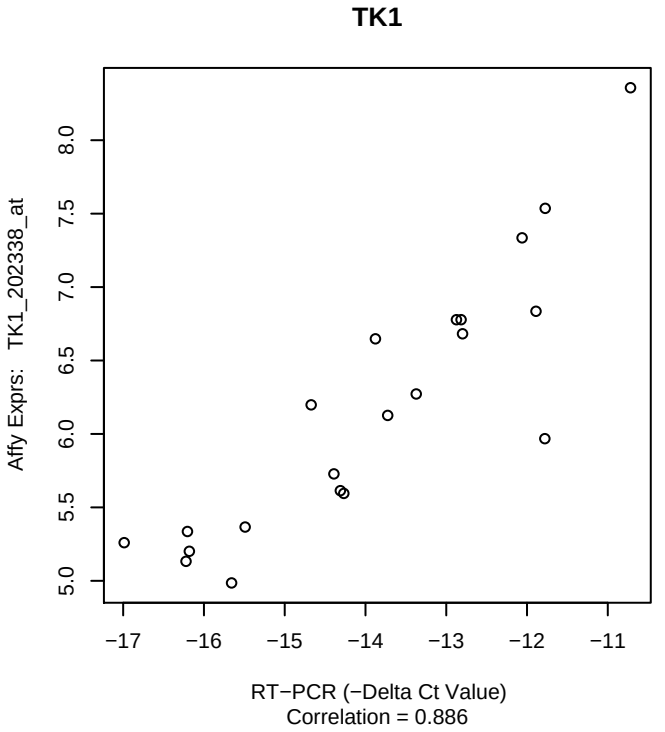

NME
Affy Exprs: NME_201577_at
RT−PCR (−Delta Ct Value)
Correlation = 0.885

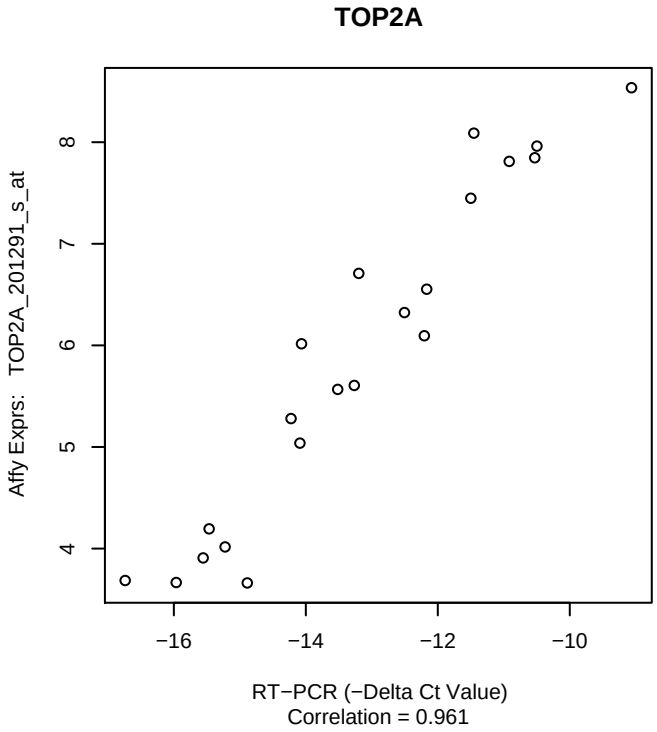

NUSAP1
Affy Exprs: NUSAP1_218039_at
RT−PCR (−Delta Ct Value)
Correlation = 0.957
PAFAH1B3
Affy Exprs: PAFAH1B3_203228_at
RT−PCR (−Delta Ct Value)
Correlation = 0.833
PBK
Affy Exprs: PBK_219148_at
RT−PCR (−Delta Ct Value)
Correlation = 0.859
PCNA
Affy Exprs: PCNA_201202_at
RT−PCR (−Delta Ct Value)
Correlation = 0.921
PRC1
Affy Exprs: PRC1_218009_s_at
RT−PCR (−Delta Ct Value)
Correlation = 0.88
RACGAP1
Affy Exprs: RACGAP1_222077_s_at
RT−PCR (−Delta Ct Value)
Correlation = −0.105

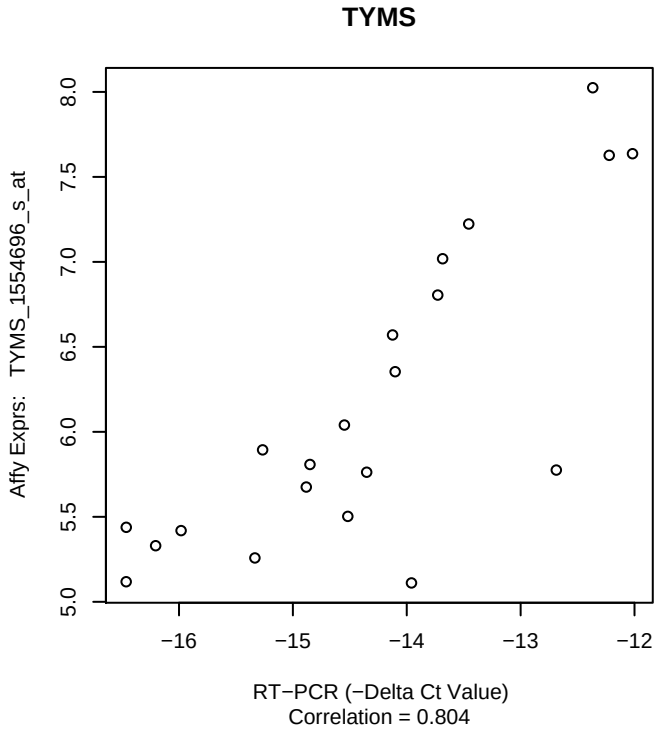

RRM2
Affy Exprs: RRM2_201890_at
RT−PCR (−Delta Ct Value)
Correlation = 0.975


SMC4
Affy Exprs: SMC4_201663_s_at
RT−PCR (−Delta Ct Value)
Correlation = 0.562


SQLE
Affy Exprs: SQLE_209218_at
RT−PCR (−Delta Ct Value)
Correlation = 0.938


TK1
Affy Exprs: TK1_202338_at
RT−PCR (−Delta Ct Value)
Correlation = 0.886


TOP2A
Affy Exprs: TOP2A_201291_s_at
RT−PCR (−Delta Ct Value)
Correlation = 0.961


TYMS
Affy Exprs: TYMS_1554696_s_at
RT−PCR (−Delta Ct Value)
Correlation = 0.804