

## Supplementary data for

### Study of circRNAs from plasma extracellular vesicle-enriched samples for early-stage non-small cell lung cancer detection using a multiplex hybridization-based method

Carlos Pedraz-Valdunciel, Stavros Giannoukakos, Ana Gimenez-Capitán, Diogo Fortunato, Martyna Filipiska, Jordi Bertrán-Alamillo, Jillian W.P. Bracht, Joselyn Valarezo, Natasa Zarovni, Alberto Fernández-Hilario, Michael Hackenberg, Andrés Aguilar-Hernández, Miguel Angel Molina-Vila, Rafael Rosell

Correspondence to: cpedraz@panoncology.com / carlospedraz@icloud.com / rrosell@iconcologia.net

#### Supplementary Tables

**Table S1.** CircRNAs detected in the different plasma volumes of the same patient with 14 and 20 cycles of pre-amplification.

		500 µL				1000 µL				1500 µL			
		14 cycles		20 cycles		14 cycles		20 cycles		14 cycles		20 cycles	
		R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2
Detected circRNA (counts > 0)		31	24	38	28	29	25	15	16	8	15	16	25
Detected circRNA (counts > 10)		20	18	33	24	9	5	4	0	0	1	1	5

*R= replica*

**Table S2.** CircRNAs detected in the plasma of the same individual subjected to 10, 12 and 14 pre-amplification cycles.

	500 µL					
	10 cycles		12 cycles		14 cycles	
	R1	R2	R1	R2	R1	R2
Detected circRNA (counts > 0)	58	45	58	47	71	47
Detected circRNA (counts > 10)	38	32	39	41	48	40

*R= replica*

**Table S3.** circRNAs identified in early-state NSCLC and non-cancer control cohorts.

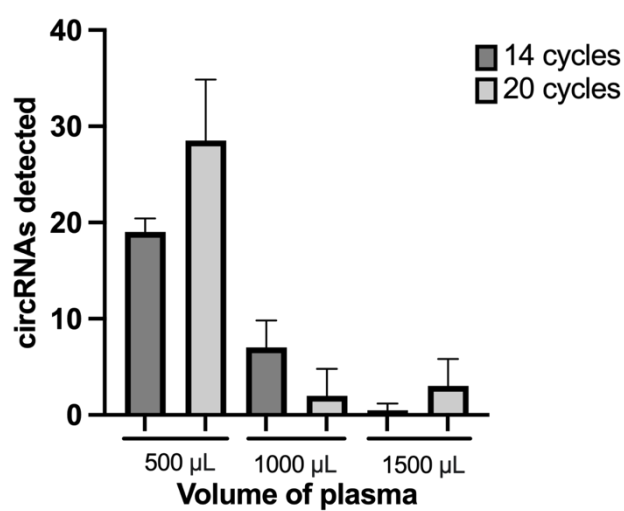
circRNAs expressed only in the NSCLC cohort	circRNAs expressed in NSCLC and control cohorts	circRNAs expressed only in the control cohort
<i>circTMEM39B</i> , <i>circZFR</i> , <i>circRHD</i> , <i>circPIK3C2B</i>	<i>circHOMER1</i> , <i>circFARSA</i> , <i>circHIPK3</i> , <i>circRUSC2</i> , <i>circC1orf116</i> , <i>circAHNAK</i> , <i>circBACH2</i> , <i>circEPB41L2</i> , <i>circSEMA5A</i> , <i>circCHD1L</i> , <i>circPIK3R1</i> , <i>circADAM22</i> , <i>circDENND1B</i> , <i>circFUT8</i> , <i>circDUS2L</i> , <i>circRANGAP1</i> , <i>circPDE5A</i> , <i>circSNX25</i> , <i>circNUPL2</i> , <i>circCSPP1</i> , <i>circCOL11A1</i> , <i>circNEDD4L</i> , <i>circNEDD4L-2</i> , <i>circUHRF1</i> , <i>circTASP1</i> , <i>circAASDH</i> , <i>circMYBL1</i> , <i>circB4GALT2</i> , <i>circCHST15</i> , <i>circCORO1C</i> , <i>circVRK1</i> , <i>circMGA</i> , <i>circGAS8</i> , <i>circSMAD2</i> , <i>circSLC8A1</i> , <i>circPMS1</i> , <i>circCCDC134</i> , <i>circFOXPI</i> , <i>circUBXN7</i> , <i>circSMARCA5</i> , <i>circCCNB1</i> , <i>circFAM13B</i> , <i>circC1GALT1</i> , <i>circRUNX1</i> , <i>circLIN54</i> , <i>circSND1</i> , <i>circCLK1</i> , <i>circHIBADH</i> , <i>circCHN1</i> , <i>circLYPLAL1</i> , <i>circPSD3</i> , <i>circSOX13</i> , <i>circRDH11</i> , <i>circYWHAZ</i> , <i>circDNA2</i> , <i>circANXA7</i> , <i>circZCCHC6</i> , <i>circTXNDC11</i> , <i>circDHCR24</i> , <i>circACP6</i> , <i>circUSP3</i> , <i>circCHD2</i> , <i>circITGAX</i> , <i>circBANP</i> , <i>circACACA</i> , <i>circBNC2</i>	<i>circNOL6</i> , <i>circNUP98</i>

**Table S4.** Normalized counts of differentially expressed circRNAs found in the early-stage NSCLC cohort.

circRNA	Gene	Mean non-cancer cohort	Mean early-stage NSCLC cohort (controls)	t-test	Fold change
circRNA_HIPK3	HIPK3	161	361	0.042	2.24
hsa_circRNA_001640	EPB41L2	109	230	0.002	2.11
hsa_circRNA_100421	DENN1B	27	62	0.049	2.31
hsa_circRNA_103809	ZFR	0.41	4	0.014	10.98
hsa_circ_0001495	CCNB1	7	21	0.040	3.10
hsa_circ_0001675	C1GALT1	1	7	0.009	10.58
hsa_circ_0007037	ZCCHC6	54	138	0.039	2.57
hsa_circ_0035654	USP3	3	22	0.016	7.13

NSCLC= non-small cell lung cancer

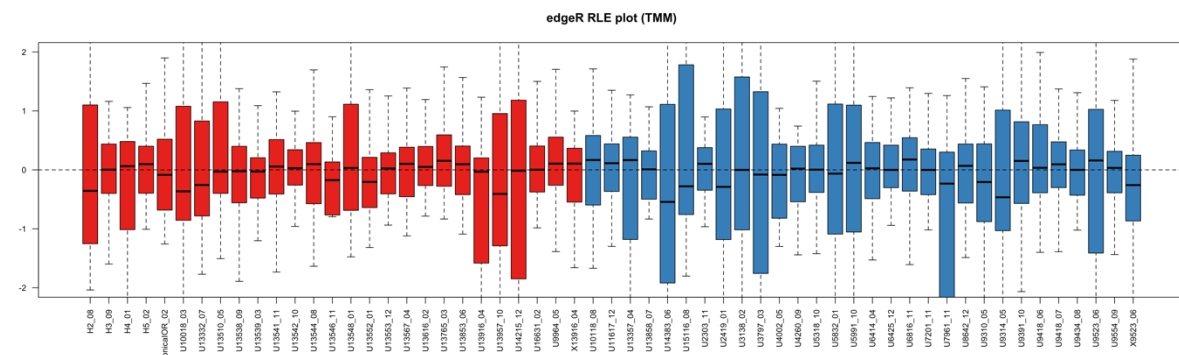
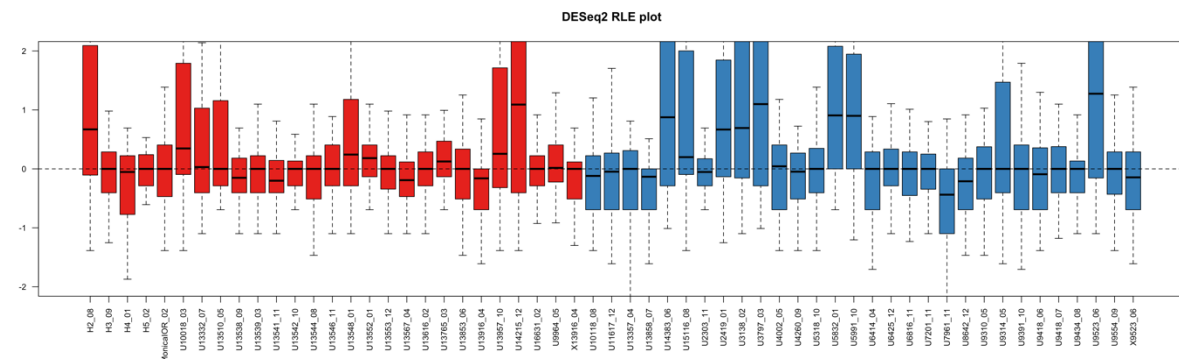
### Supplementary Figures



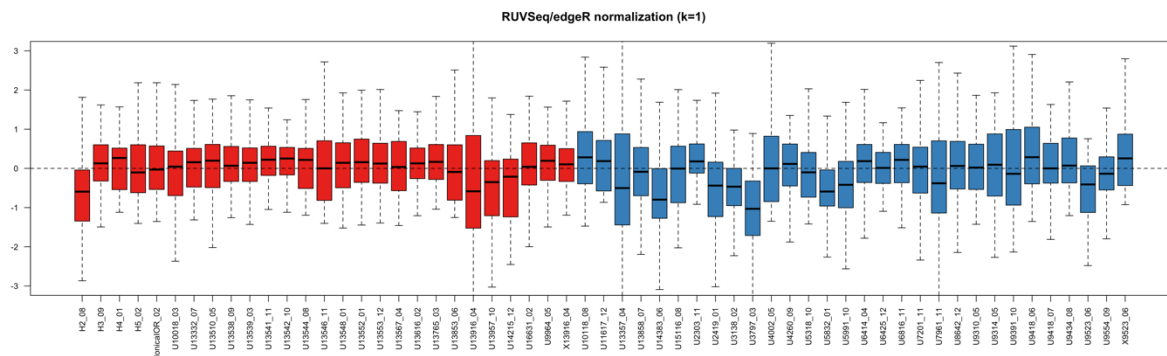
**Figure S1.** Plasma input testing. Number of circRNAs with a score > 10 counts after background removal for each of the volumes tested. Error bars indicate standard deviation.

[illegible]

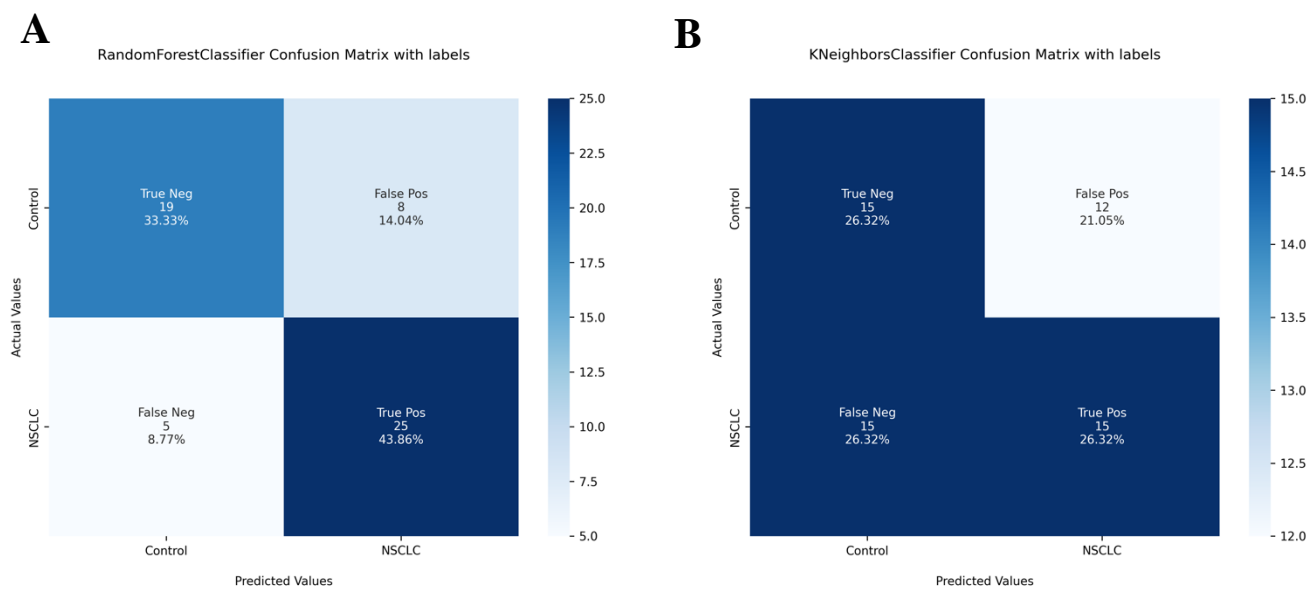
**Figure S2.** Principal Component Analysis of the transformed raw data. Outliers highlighted by interquartile range plot analysis were also found on the far-left side of the plot.



C



**Figure S3.** Assessment of the different normalization processes by RLE plot analysis. **(A)** RLE plot of normalized counts by DESeq2. **(B)** RLE plot of normalized counts by edgeR. **(C)** RLE plot of normalized counts using a combination of the edgeR-RUVg methods ( $k = 1$ ).



**Figure S4.** Confusion matrices summarizing the performance of the different classification algorithms. **(A)** Confusion matrix of the of the RFF classifier. **(B)** Confusion matrix of the KNN classifier. In both cases, 5-Fold CV was used.

*RF = Random Forest, KNN = K-Nearest Neighbor, 5-CV = 5-Cross Validation, RFE = Recursive Feature Elimination.*