

# Supplementary Material: Integration of miRNA-seq and Clinical data to Identify Breast Cancer Subtype specific miRNAs using Survival Analysis and their Interactions with Genes and Transcription Factors

## 1 SUPPLEMENTARY TABLES AND FIGURES

### 1.1 Tables

**Table S1.** List of top 100 miRNAs for each breast cancer subtype after performing survival analysis. (Please see Excel file named 'Top100-miRNAs')

**Table S2.** The parameters of seven different classifiers

Logistic Regression		Support Vector Machine		Decision Tree	
Parameter	Value	Parameter	Value	Parameter	Value
class_weight	balanced	class_weight	balanced	class_weight	balanced
solver	lbfgs	Random Forest		Multilayer Perceptron	
max_iter	200			Parameter	Value
multi_class	multinomial	class_weight	balanced	max_iter	1000
KNN		Naïve Bayes		solver	lbfgs
				Parameter	Value
n_neighbors	3	var_smoothing	10 <sup>-9</sup>		

**Table S3.** Statistics of the samples of breast cancer stage for different subtypes. For the 1-star miRNAs only the patients with respective cancer subtype are used, while in case of the 4-star miRNAs all patients with stage data are used

miRNA set	Breast cancer stage			
	I	II	III	IV
4-star	46	107	29	2
1-star LA	30	41	12	1
1-star LB	5	25	8	1
1-star H2	3	13	6	0
1-star BL	8	28	3	0

**Table S4.** Results of Anova test on the stage-wise data for 4-star and 1-star miRNAs (Please see Excel file named 'Anova-Stage')

**Table S5.** Count of miRNAs, genes, and transcription factors (TFs) after specific steps of the filtering procedure

miRNAs	Steps	miRNA	Gene	TF
4-star	1, 2	44	9300	126
	3, 4, 5	36	8863	699
	6, 7, 8	36	1027	93
	9, 10, 11	21	80	15
1-star LA	1, 2	12	4447	558
	3, 4, 5	8	3911	44
	6, 7, 8	8	405	35
	9, 10, 11	4	10	4
1-star LB	1, 2	14	3409	551
	3, 4, 5	12	3164	50
	6, 7, 8	12	316	36
	9, 10, 11	9	17	7
1-star HER2-E	1, 2	9	4204	570
	3, 4, 5	7	3958	47
	6, 7, 8	7	411	38
	9, 10, 11	4	46	6
1-star BL	1, 2	15	4500	595
	3, 4, 5	14	4500	35
	6, 7, 8	14	344	34
	9, 10, 11	7	41	4

**Table S6.** The detailed results of interactions of miRNA→Gene, TF→Gene, and TF→miRNA for 4-star and 1-star miRNAs as well as for the combined networks

miRNA set	File
Combined network	Excel file named 'Combined_miRNA_Query_details'
4-star	Excel file named '44_miRNA_Query_details'
1-star LA	Excel file named 'LA_miRNA_Query_details'
1-star LB	Excel file named 'LB_miRNA_Query_details'
1-star HER2-E	Excel file named 'H2_miRNA_Query_details'
1-star BL	Excel file named 'BL_miRNA_Query_details'

**Table S7.** The Transcription factors identified as the result of filtering procedure, in each for 4-star and 1-star miRNA sets, and the common set (Please see the Excel file named 'TF-Common')

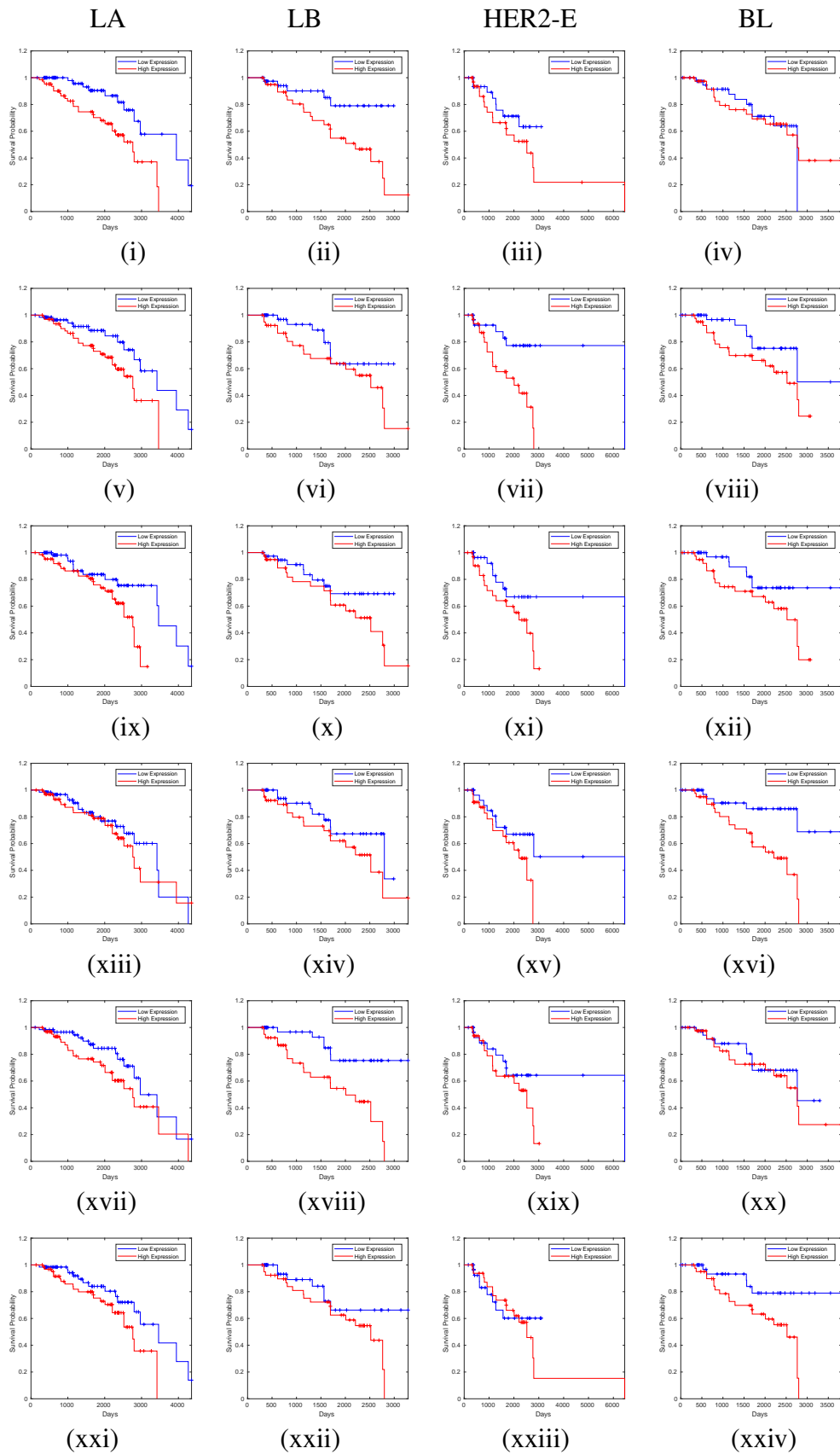
**Table S8.** The detailed results of PPI interaction analysis for 4-star and 1-star miRNAs (Please see the Excel file named 'PPI-Interaction')

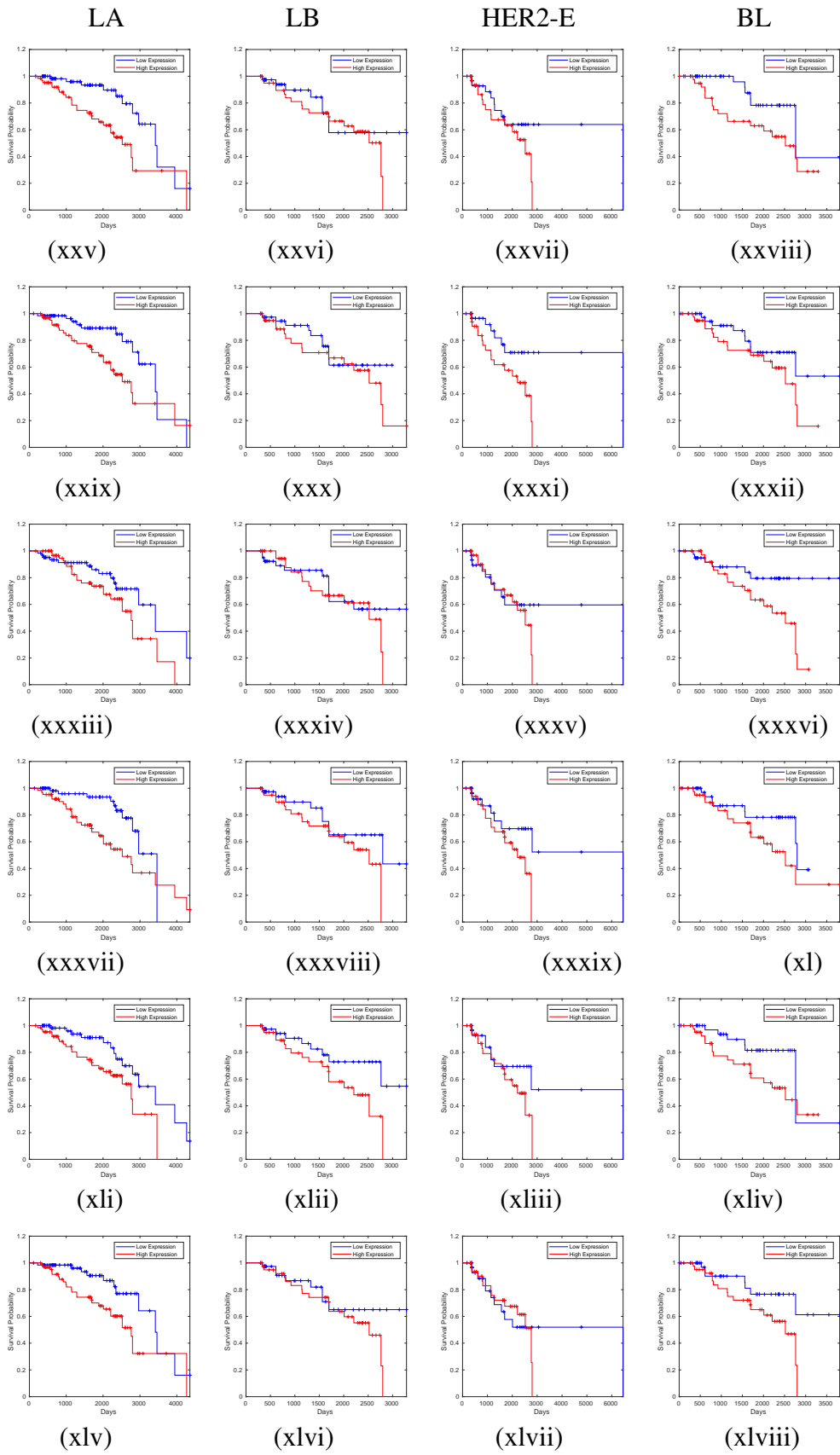
**Table S9.** The detailed results of KEGG Pathway Analysis, from gene perspective, for 4-star and 1-star miRNAs (Please see the Excel file named 'KEGG')

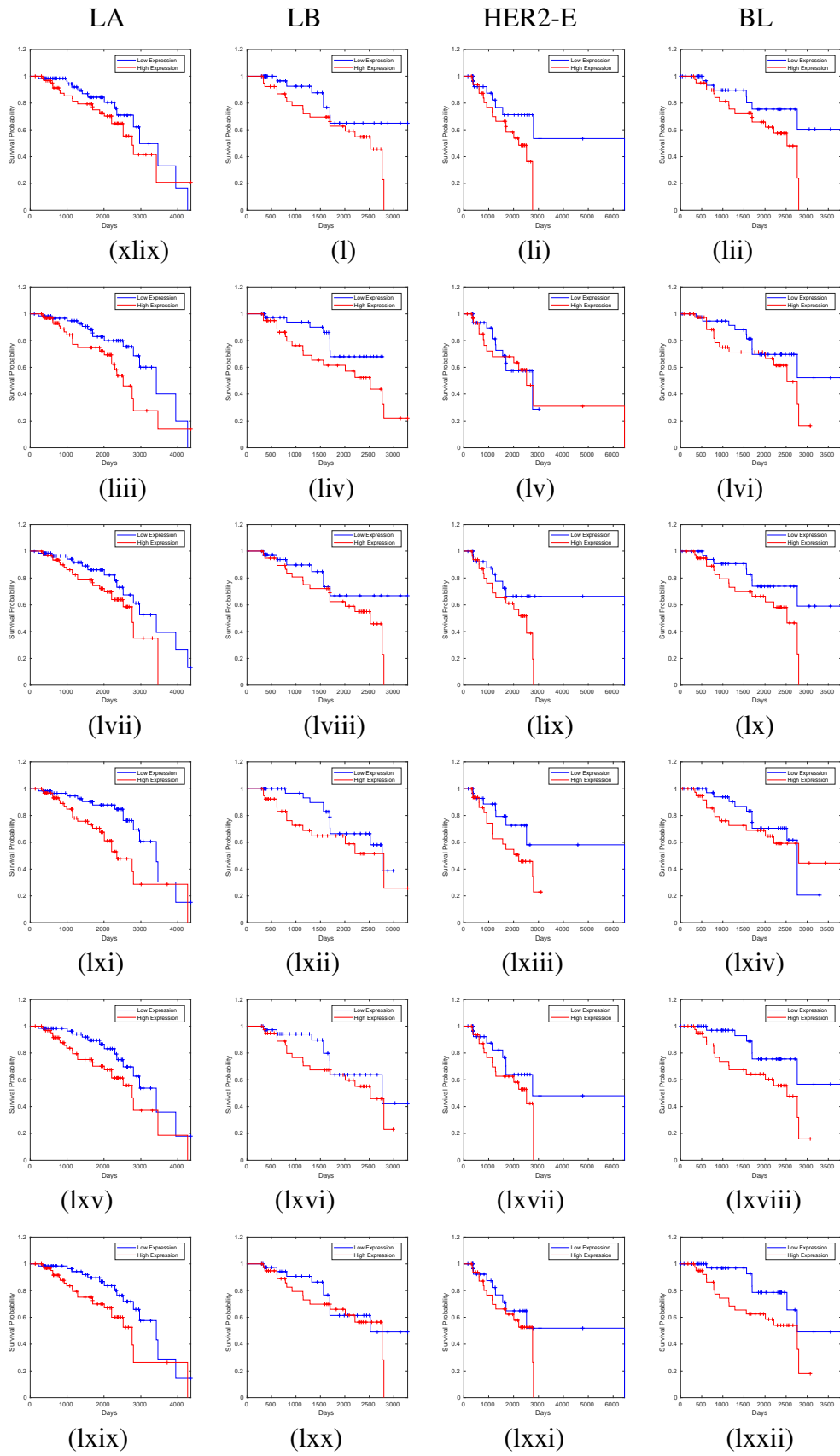
**Table S10.** The detailed results of GO Enrichment Analysis, from gene perspective, for 4-star and 1-star miRNAs

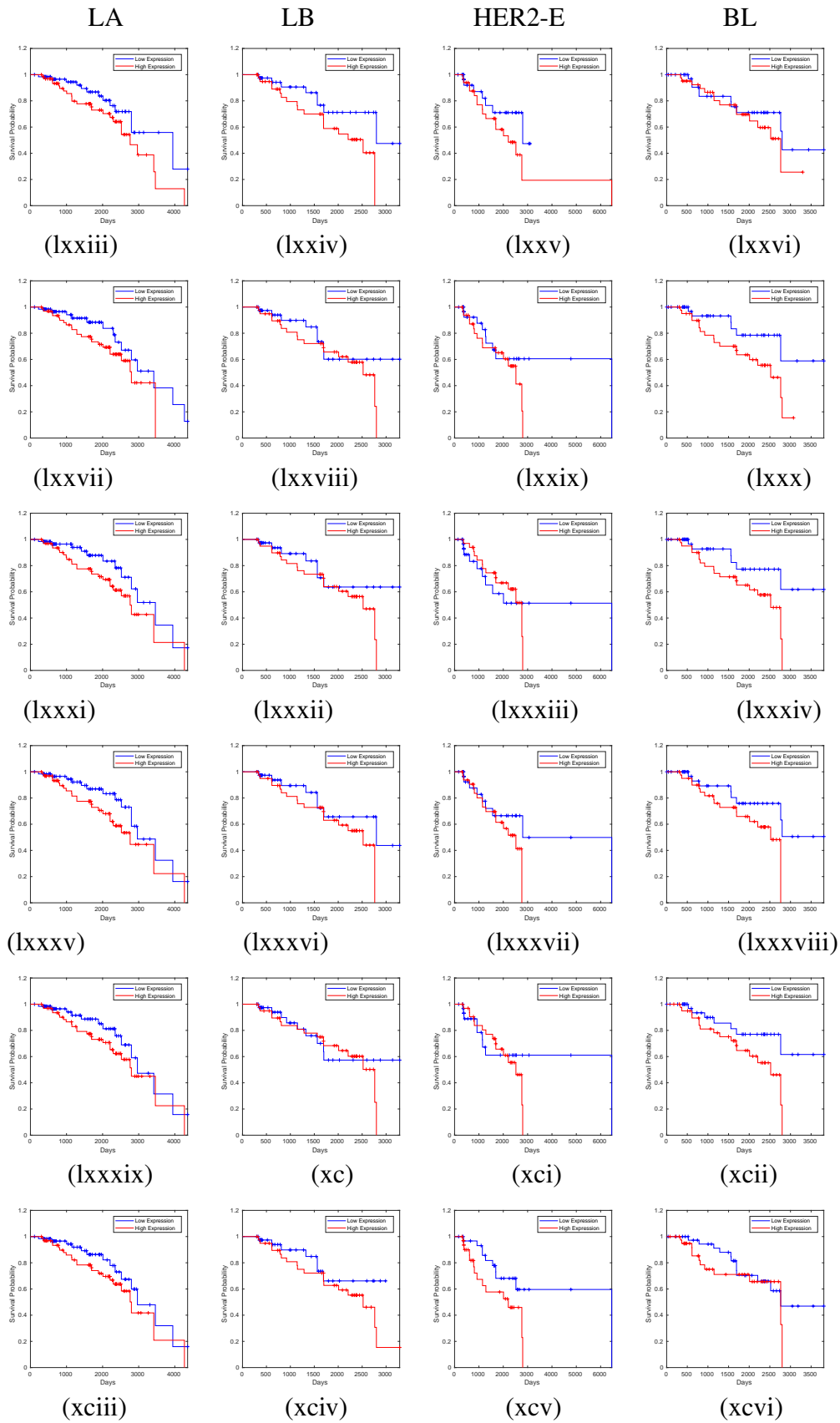
miRNAs	File
4-star	Excel file named 'GO-44'
1-star LA	Excel file named 'GO-LA'
1-star LB	Excel file named 'GO-LB'
1-star HER2-E	Excel file named 'GO-HER2-E'
1-star BL	Excel file named 'GO-BL'

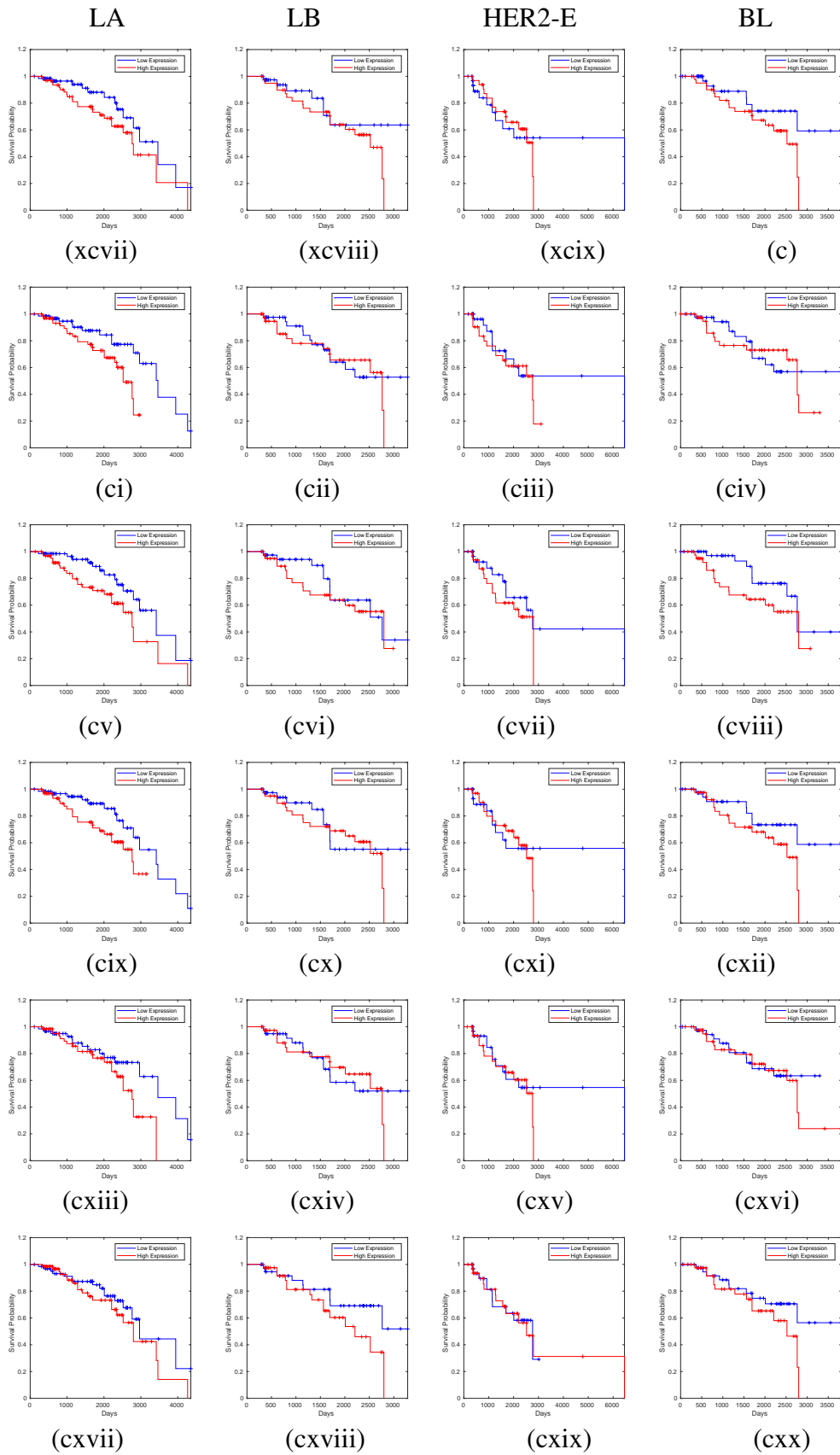
## 1.2 Figures

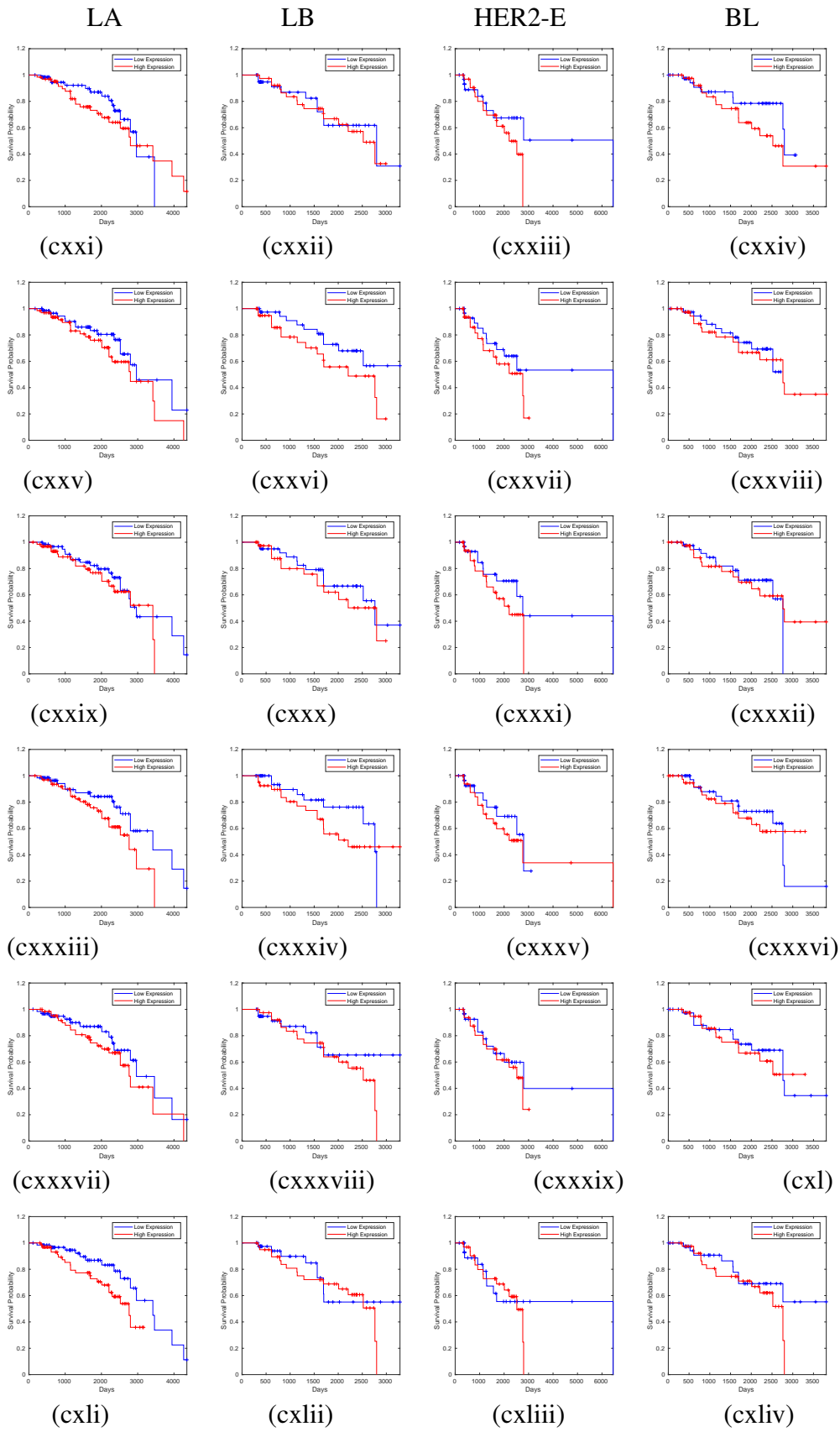














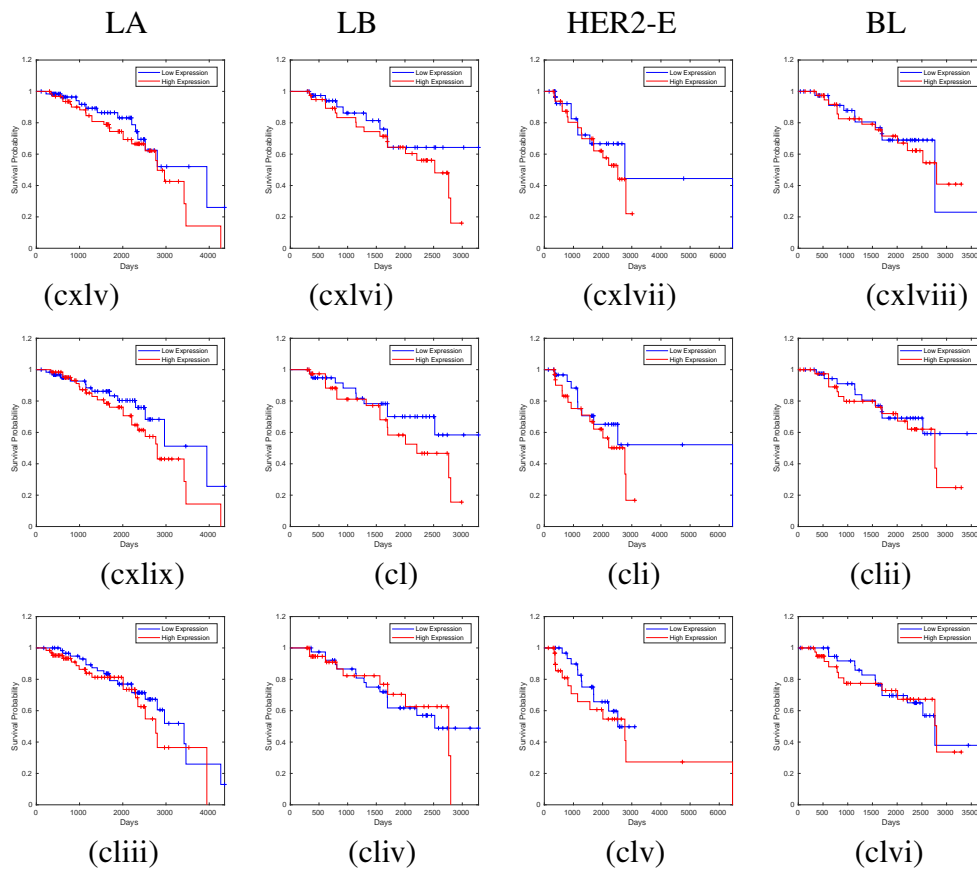
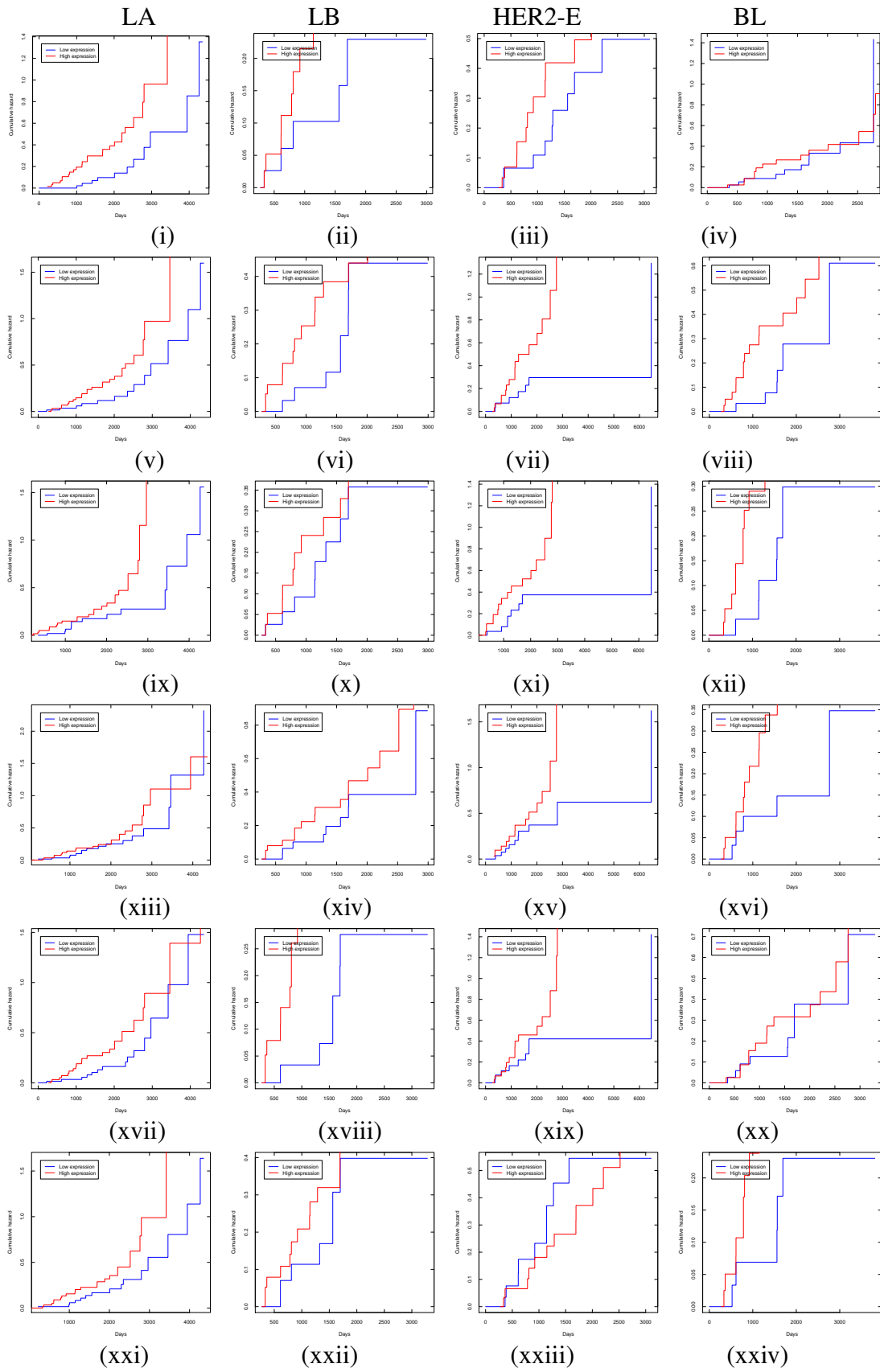
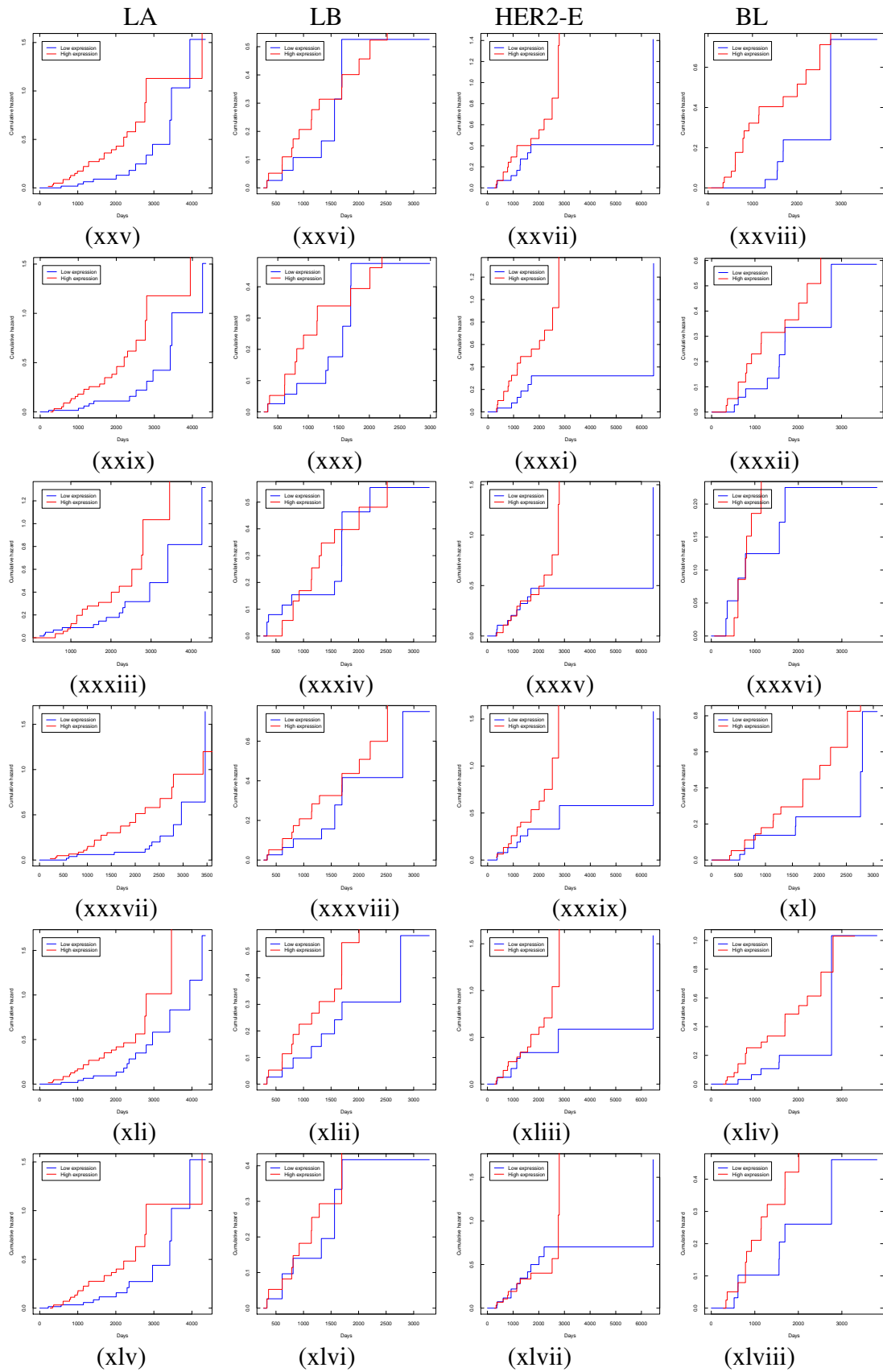
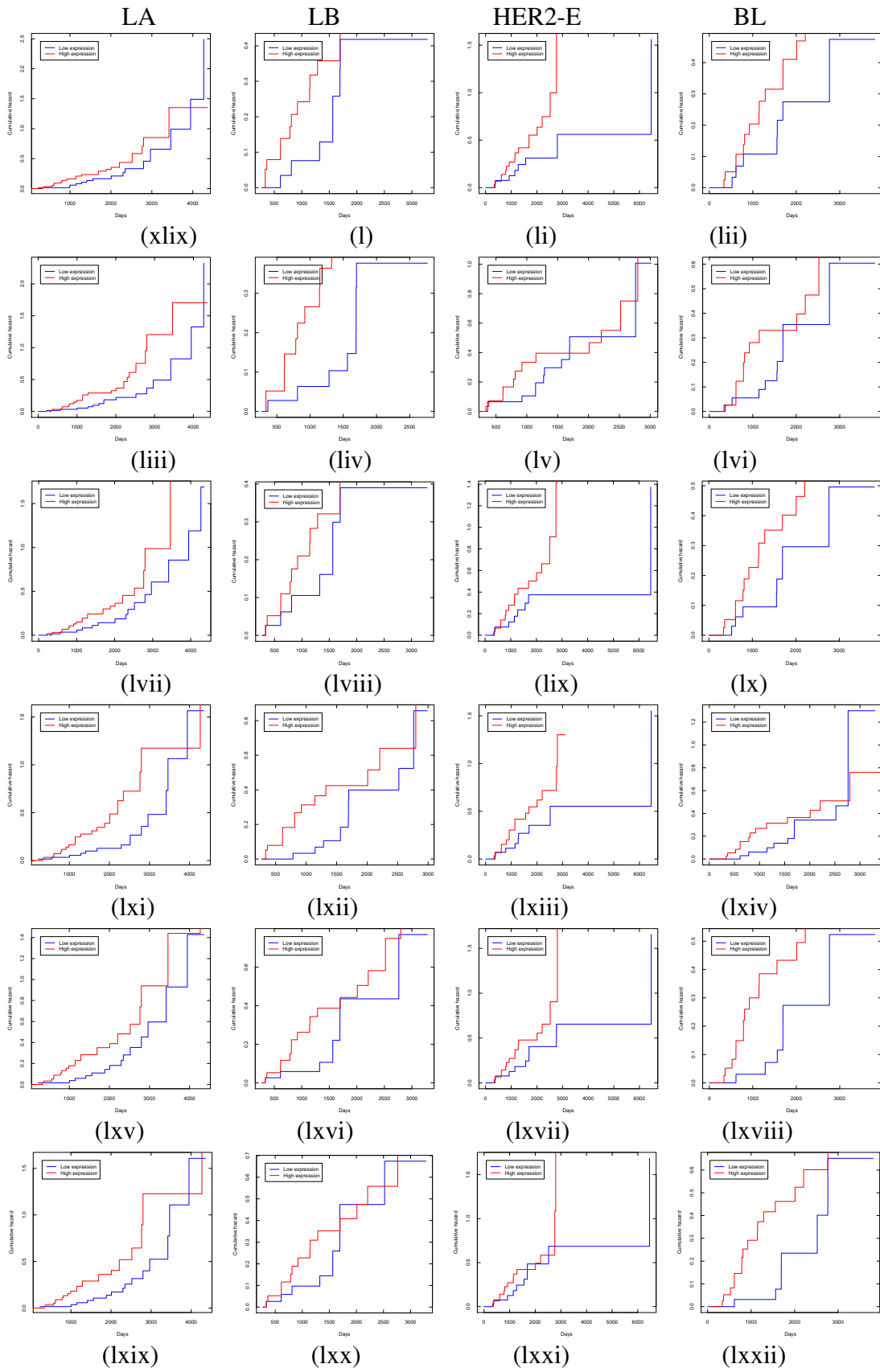
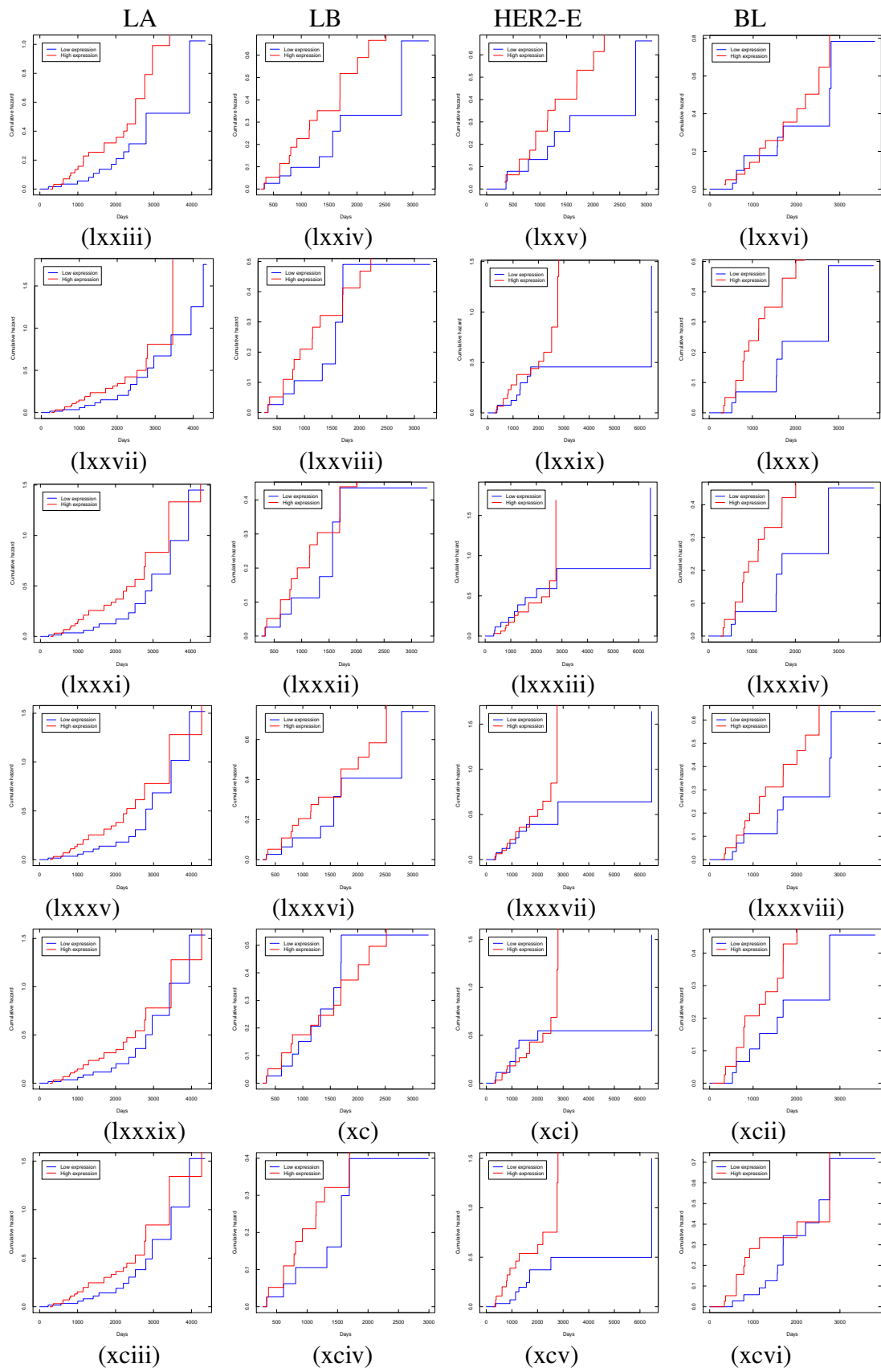


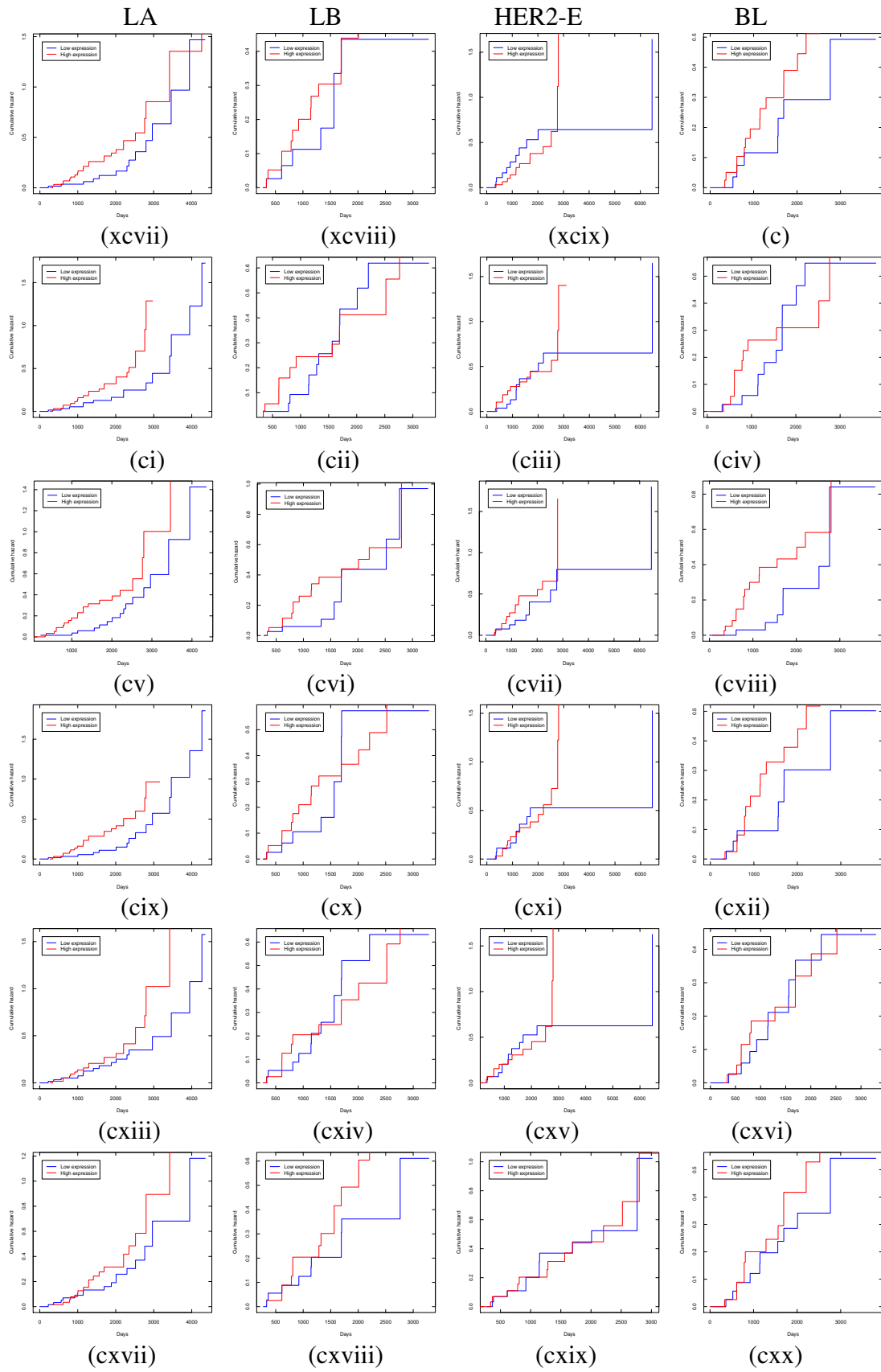
Figure S1: Survival plots for 4-star miRNAs in LA, LB, HER2-E and BL as (i)-(iv):hsa-miR-224-5p, (v)-(viii):hsa-miR-335-3p, (ix)-(xii):hsa-miR-326, (xiii)-(xvi):hsa-miR-10a-5p, (xvii)-(xx):hsa-miR-217, (xxi)-(xxiv):hsa-miR-10b-3p, (xxv)-(xxviii):hsa-miR-378a-5p, (xxix)-(xxxii):hsa-miR-193a-5p, (xxxiii)-(xxxvi):hsa-miR-664a-3p, (xxxvii)-(xl):hsa-miR-30c-2-3p, (xli)-(xliv):hsa-miR-511-5p, (xlv)-(xlviii):hsa-miR-143-3p, (xlix)-(lii):hsa-miR-10b-5p, (liii)-(lvi):hsa-miR-22-3p, (lvii)-(lx):hsa-miR-140-3p, (lxi)-(lxiv):hsa-miR-338-3p, (lxv)-(lxviii):hsa-miR-451a, (lxix)-(lxxii):hsa-miR-486-5p, (lxxiii)-(lxxvi):hsa-miR-28-3p, (lxxvii)-(lxxx):hsa-miR-139-5p, (lxxxii)-(lxxxiv):hsa-miR-125b-2-3p, (lxxxv)-(lxxxviii):hsa-miR-100-5p, (lxxxix)-(xcii):hsa-miR-195-5p, (xciii)-(xcvi):hsa-miR-584-5p, (xcvii)-(c):hsa-let-7c-5p, (ci)-(civ):hsa-miR-574-3p, (cv)-(cviii):hsa-miR-144-5p, (cix)-(cxii):hsa-miR-145-5p, (cxiii)-(cxvi):hsa-let-7e-3p, (cxvii)-(cxx):hsa-miR-24-1-5p, (cxxi)-(cxxiv):hsa-miR-30a-3p, (cxxv)-(cxxviii):hsa-miR-362-5p, (cxxix)-(cxxxii):hsa-miR-339-5p, (cxxxiii)-(cxxxvi):hsa-miR-361-3p, (cxxxvii)-(cxl):hsa-miR-30e-3p, (cxli)-(cxliv):hsa-miR-145-3, (cxlv)-(cxlviii):hsa-miR-29a-3p, (cxlix)-(clii):hsa-miR-34a-5p, (cliii)-(clvi):hsa-miR-193b-5p, where blue line indicates low expression group and red line indicates high expression group

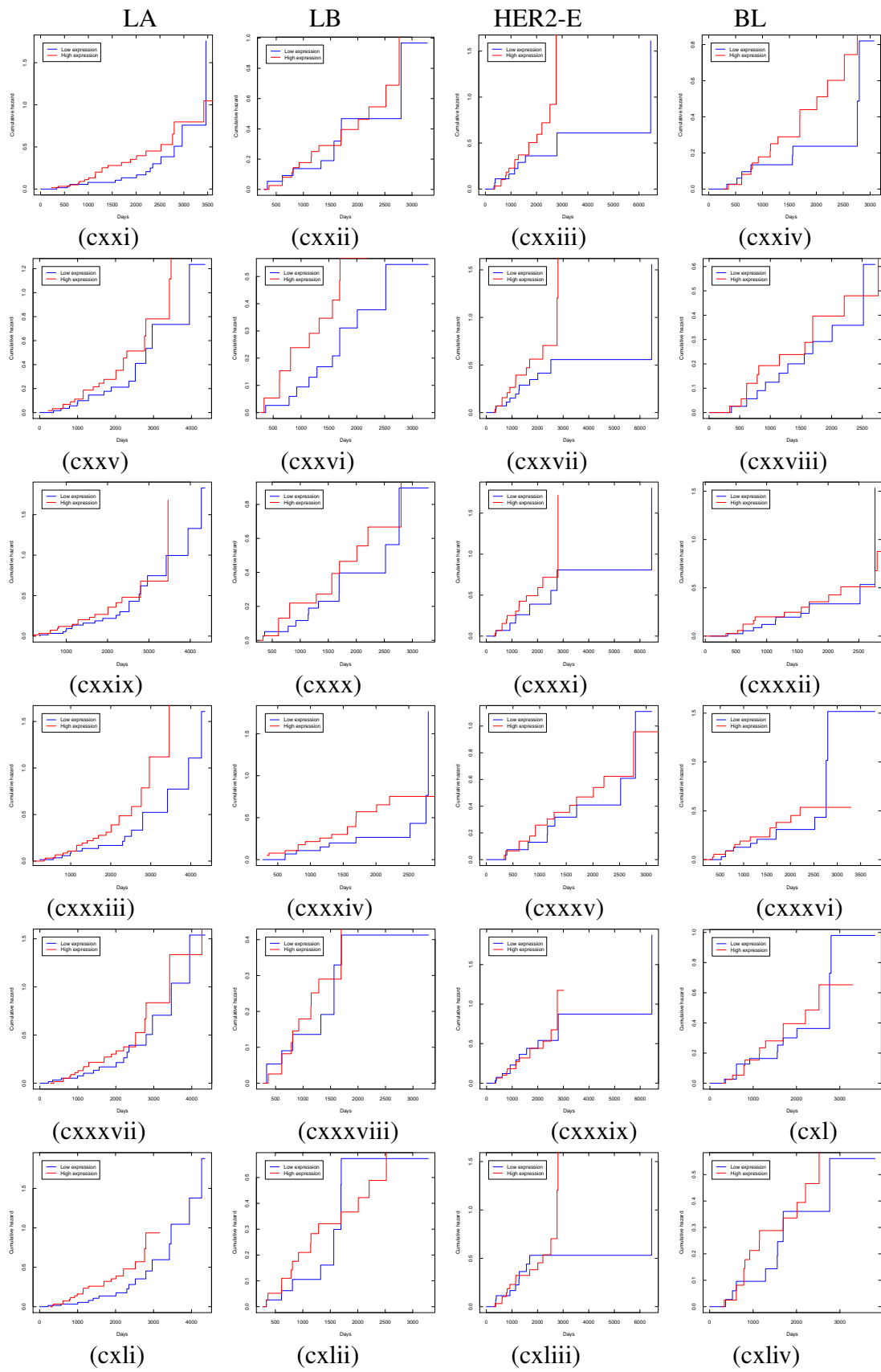












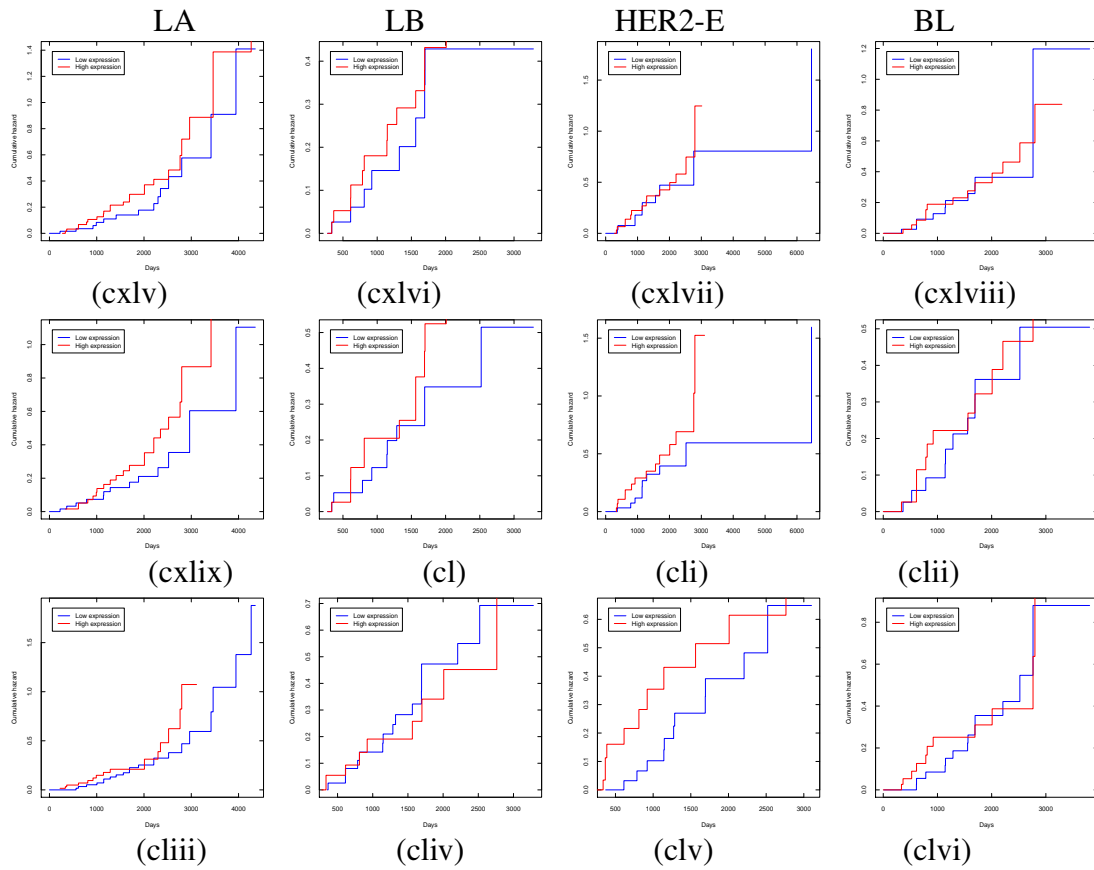


Figure S2: Cumulative hazard plots, using Nelson-Aalen estimator for 4-star miRNAs in LA, LB, HER2-E and BL as (i)-(iv):hsa-miR-224-5p, (v)-(viii):hsa-miR-335-3p, (ix)-(xii):hsa-miR-326, (xiii)-(xvi):hsa-miR-10a-5p, (xvii)-(xx):hsa-miR-217, (xxi)-(xxiv):hsa-miR-10b-3p, (xxv)-(xxviii):hsa-miR-378a-5p, (xxix)-(xxxii):hsa-miR-193a-5p, (xxxiii)-(xxxvi):hsa-miR-664a-3p, (xxxvii)-(xl):hsa-miR-30c-2-3p, (xli)-(xliv):hsa-miR-511-5p, (xlv)-(xlviii):hsa-miR-143-3p, (xlix)-(lii):hsa-miR-10b-5p, (liii)-(lvi):hsa-miR-22-3p, (lvii)-(lx):hsa-miR-140-3p, (lxi)-(lxiv):hsa-miR-338-3p, (lxv)-(lxviii):hsa-miR-451a, (lxix)-(lxxii):hsa-miR-486-5p, (lxxiii)-(lxxvi):hsa-miR-28-3p, (lxxvii)-(lxxx):hsa-miR-139-5p, (lxxxii)-(lxxxiv):hsa-miR-125b-2-3p, (lxxxv)-(lxxxviii):hsa-miR-100-5p, (lxxxix)-(xcii):hsa-miR-195-5p, (xciii)-(xcvi):hsa-miR-584-5p, (xcvii)-(c):hsa-let-7c-5p, (ci)-(civ):hsa-miR-574-3p, (cv)-(cviii):hsa-miR-144-5p, (cix)-(cxii):hsa-miR-145-5p, (cxiii)-(cxvi):hsa-let-7e-3p, (cxvii)-(cxx):hsa-miR-24-1-5p, (cxxi)-(cxxiv):hsa-miR-30a-3p, (cxxv)-(cxxviii):hsa-miR-362-5p, (cxxix)-(cxxxii):hsa-miR-339-5p, (cxxxiii)-(cxxxvi):hsa-miR-361-3p, (cxxxvii)-(cxl):hsa-miR-30e-3p, (cxli)-(cxliv):hsa-miR-145-3, (cxlv)-(cxlviii):hsa-miR-29a-3p, (cxlix)-(clii):hsa-miR-34a-5p, (cliii)-(clvi):hsa-miR-193b-5p, where blue line indicates low expression group and red line indicates high expression group



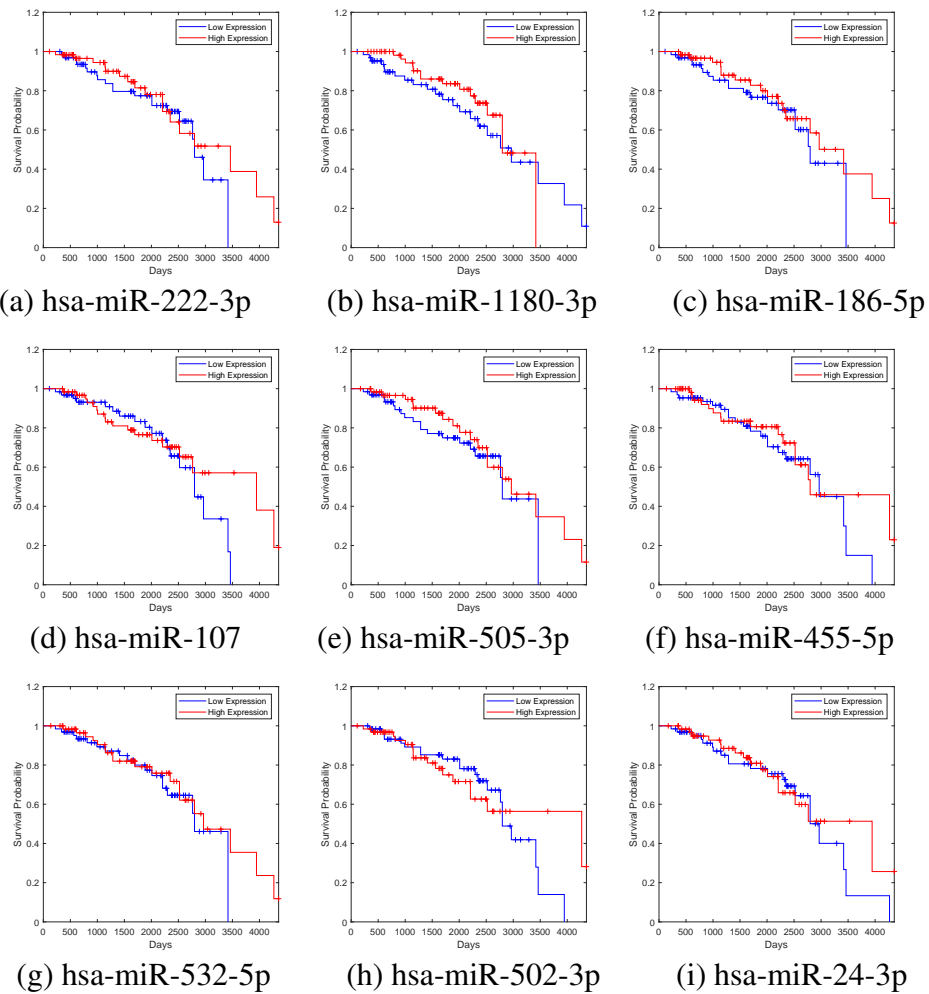


Figure S3: Survival plots for 1-star LA miRNAs where blue line indicates low expression group and red line indicates high expression group

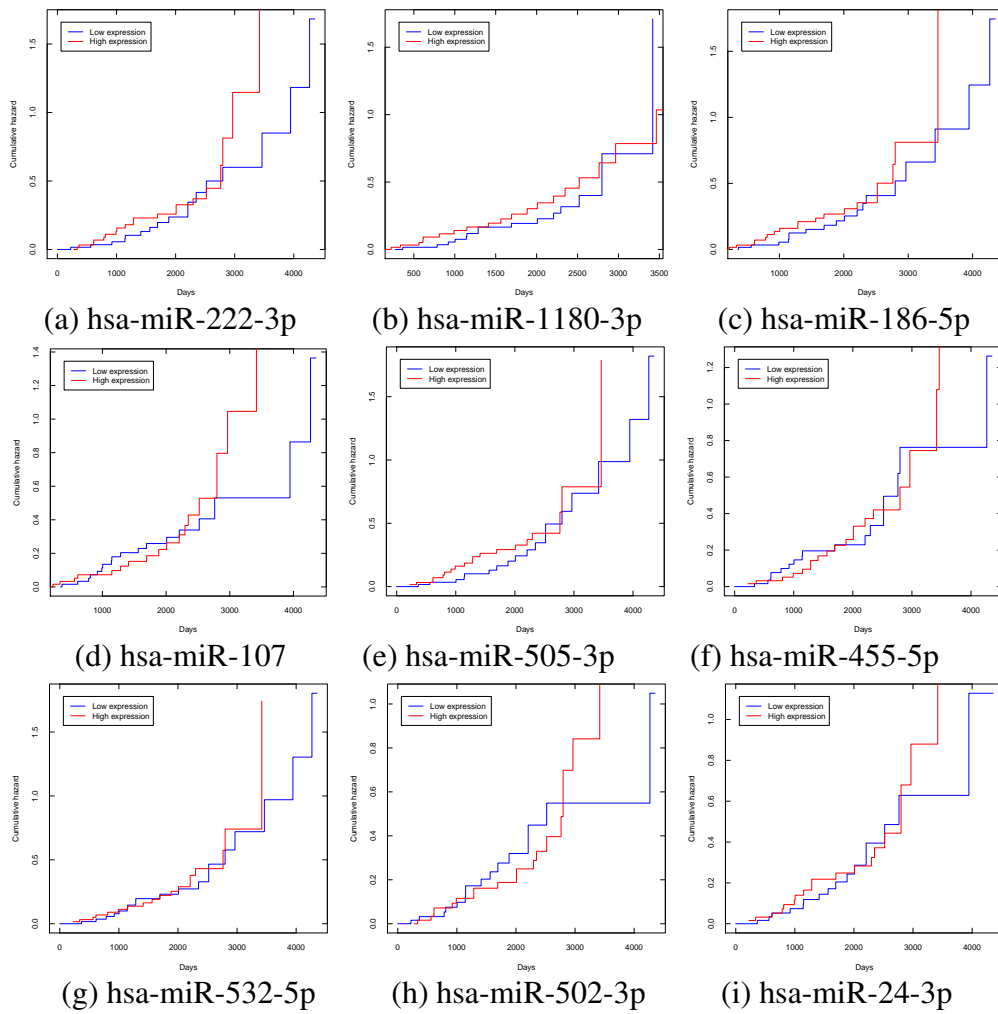


Figure S4: Cumulative hazard plots, using Nelson-Aalen estimator for 1-star LA miRNAs where blue line indicates low expression group and red line indicates high expression group

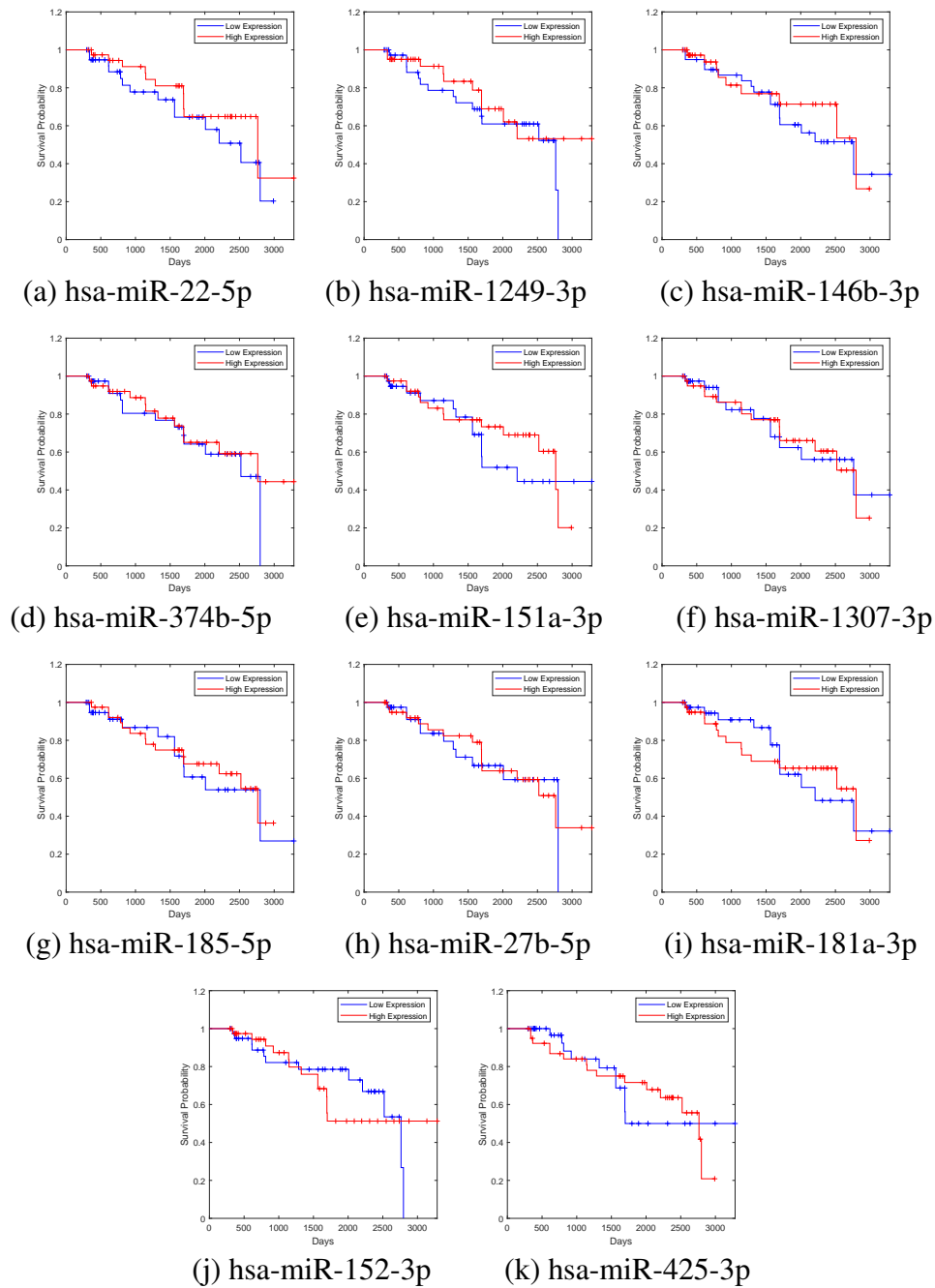


Figure S5: Survival plots for 1-star LB miRNAs where blue line indicates low expression group and red line indicates high expression group

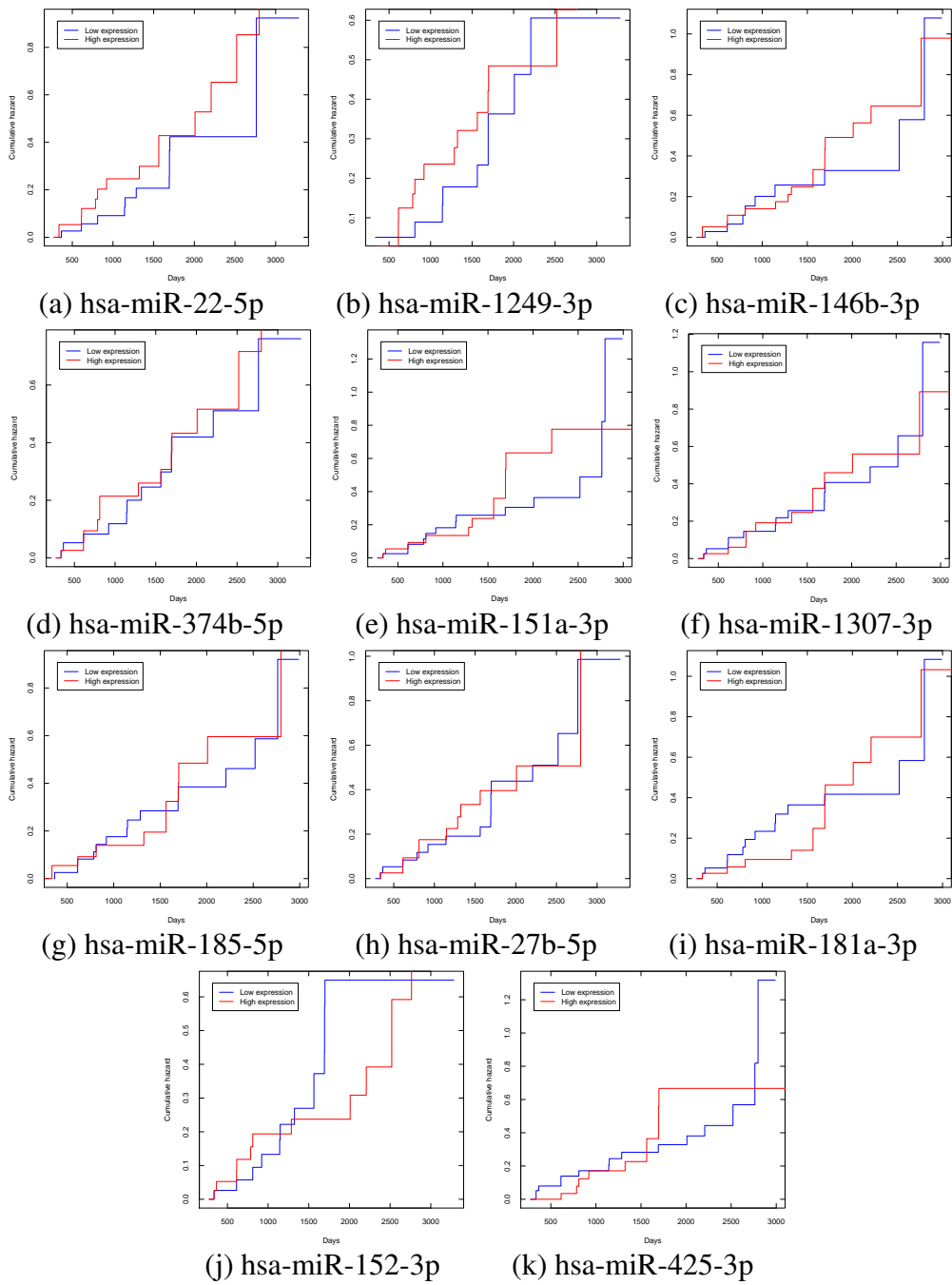


Figure S6: Cumulative hazard plots, using Nelson-Aalen estimator for 1-star LB miRNAs where blue line indicates low expression group and red line indicates high expression group

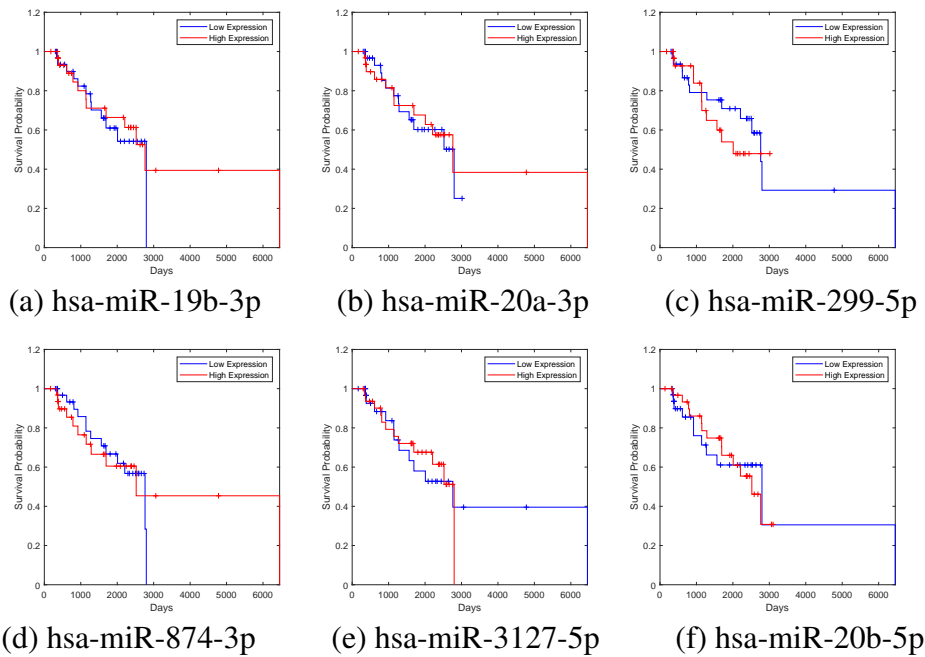


Figure S7: Survival plots for 1-star HER2-E miRNAs where blue line indicates low expression group and red line indicates high expression group

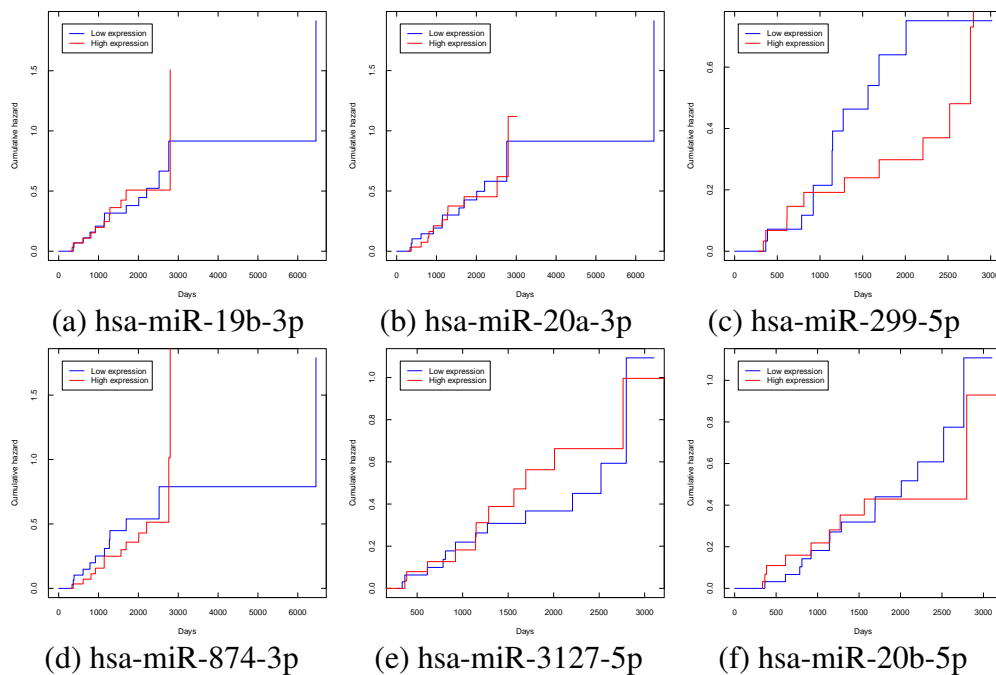


Figure S8: Cumulative hazard plots, using Nelson-Aalen estimator for 1-star HER2-E miRNAs where blue line indicates low expression group and red line indicates high expression group

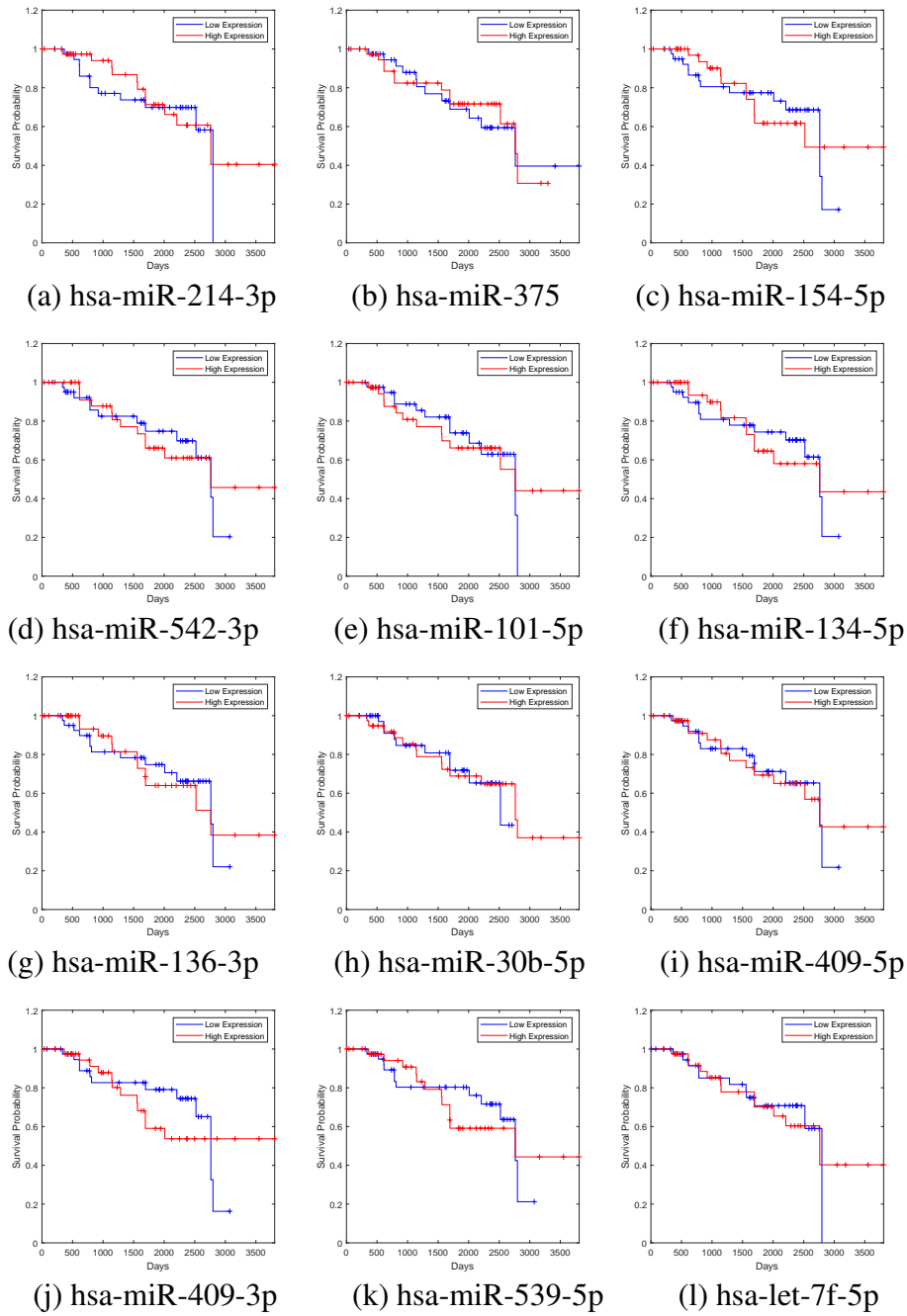


Figure S9: Survival plots for 1-star BL miRNAs where blue line indicates low expression group and red line indicates high expression group

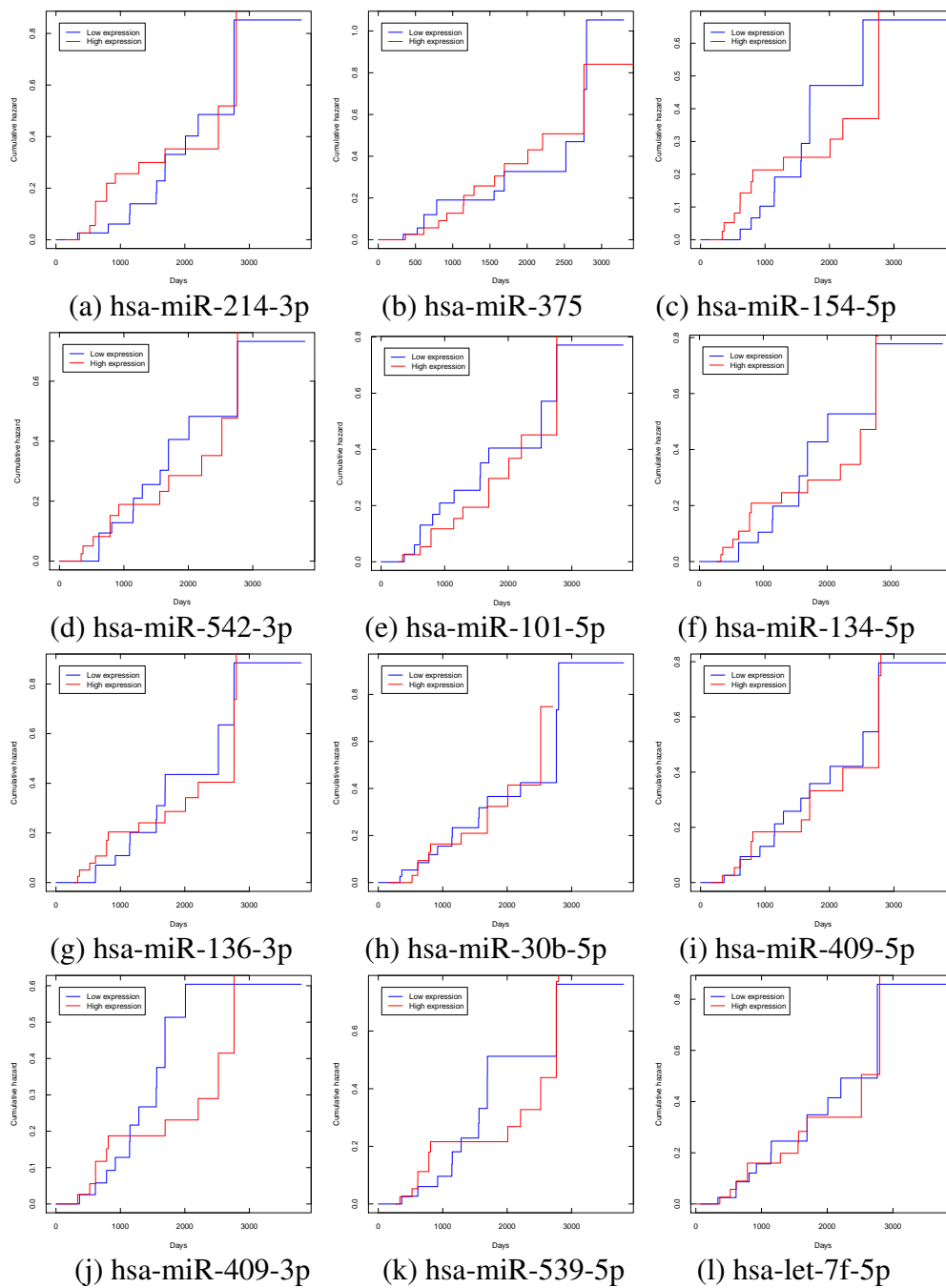


Figure S10: Cumulative hazard plots, using Nelson-Aalen estimator for 1-star BL miRNAs where blue line indicates low expression group and red line indicates high expression group

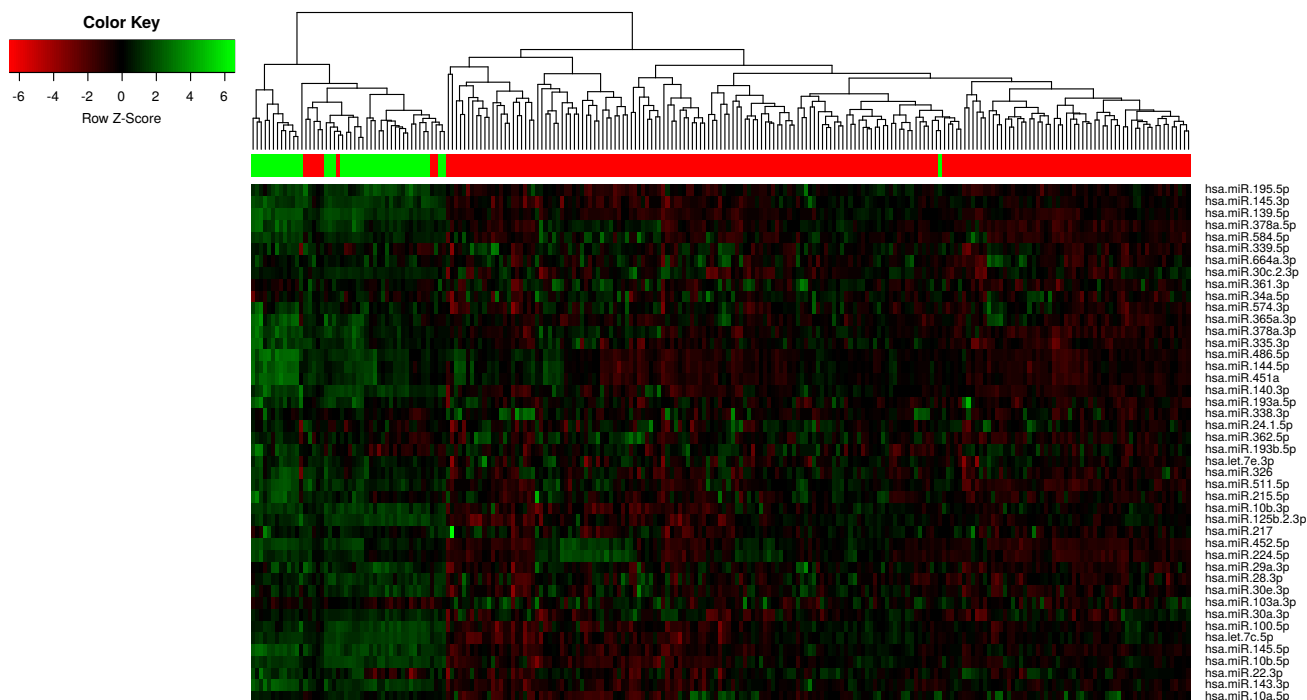
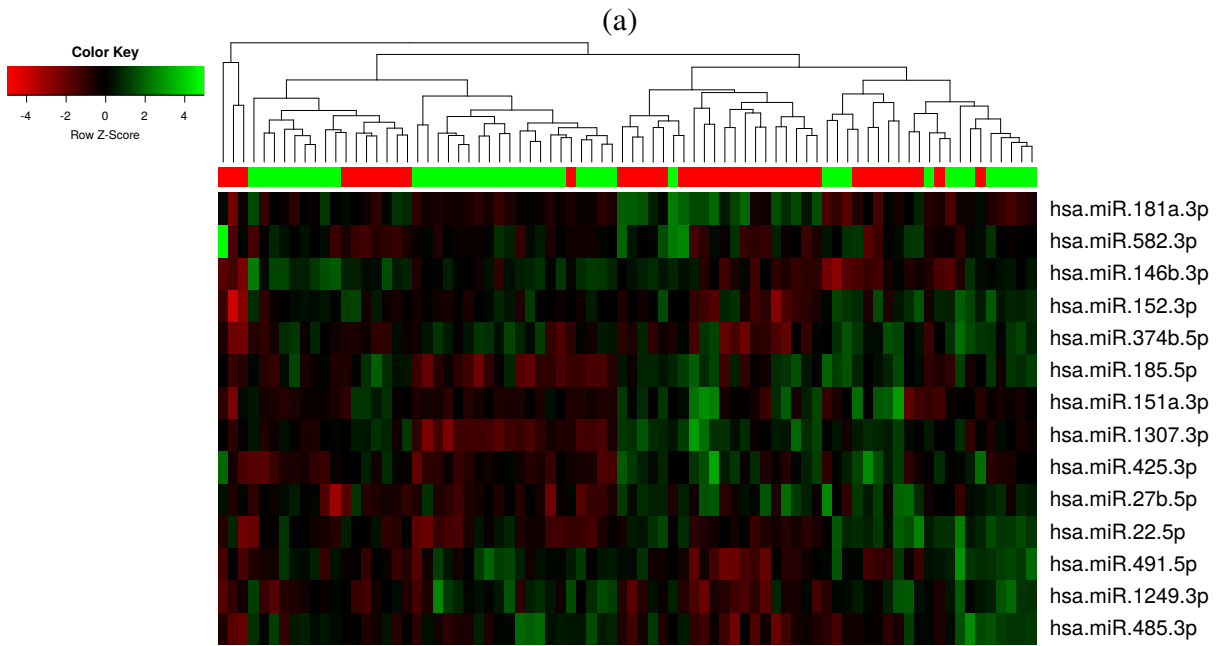
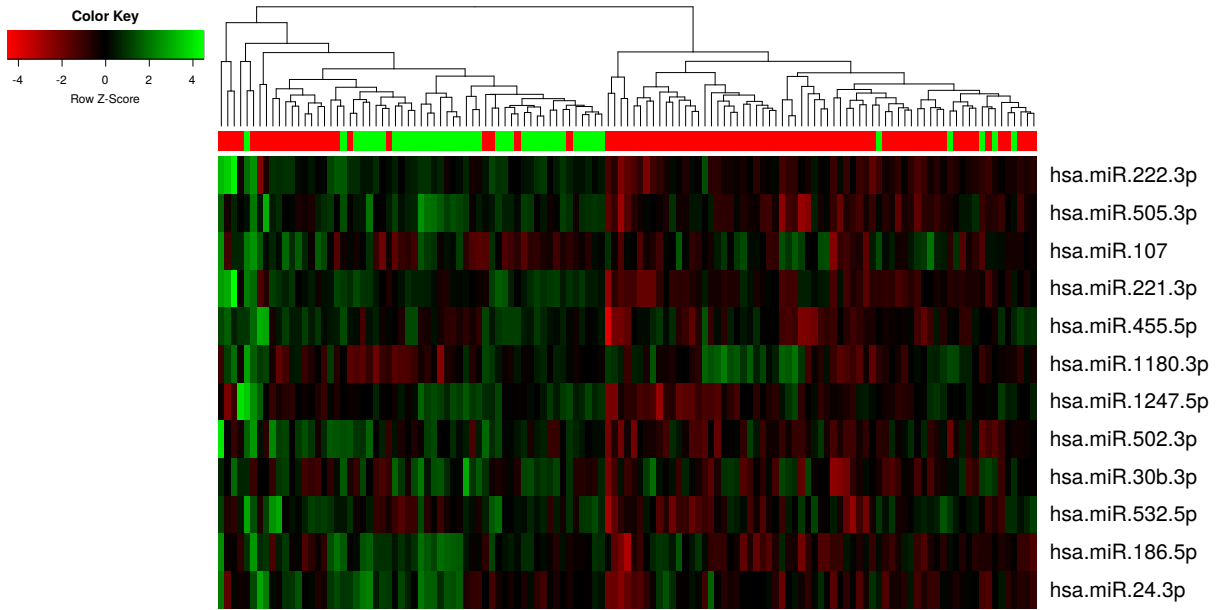
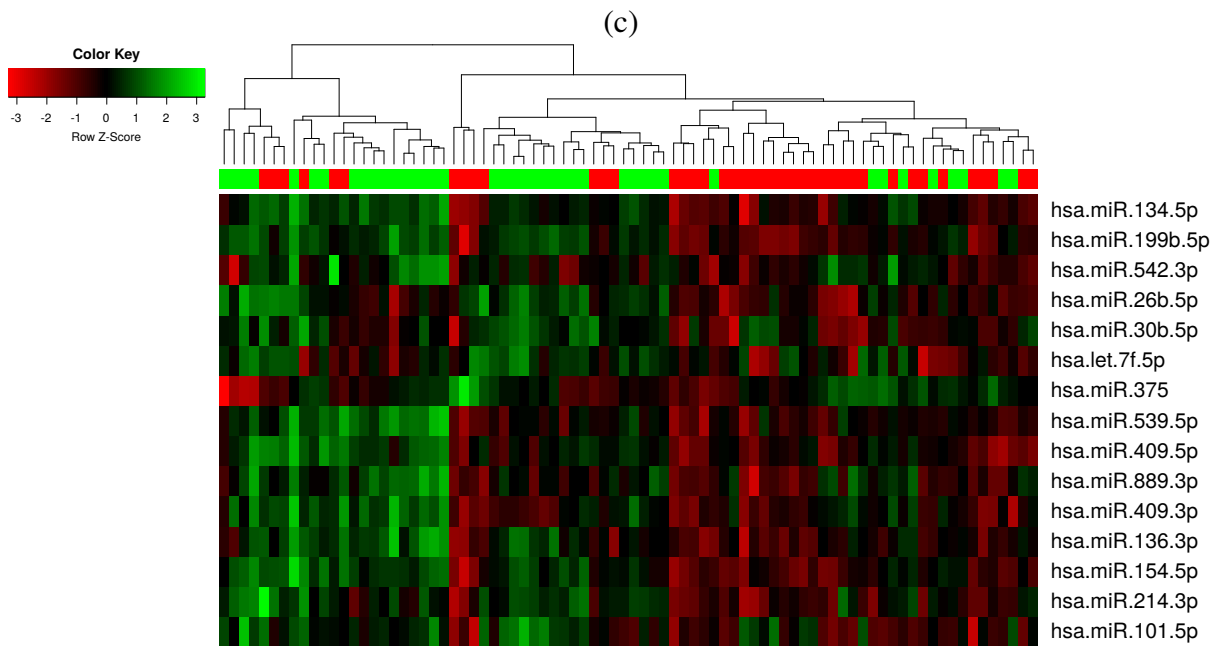
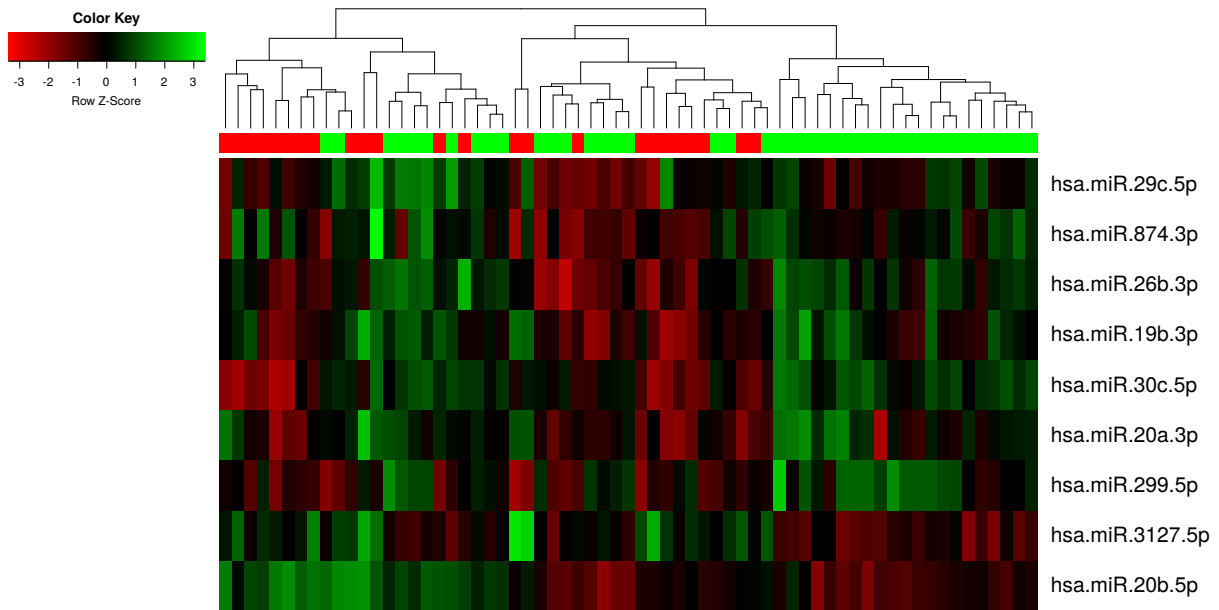


Figure S11: Clustering analysis of the 4-star miRNAs of all patients with control. Color of the top bar represents group: red for cancer (any subtype), green for control



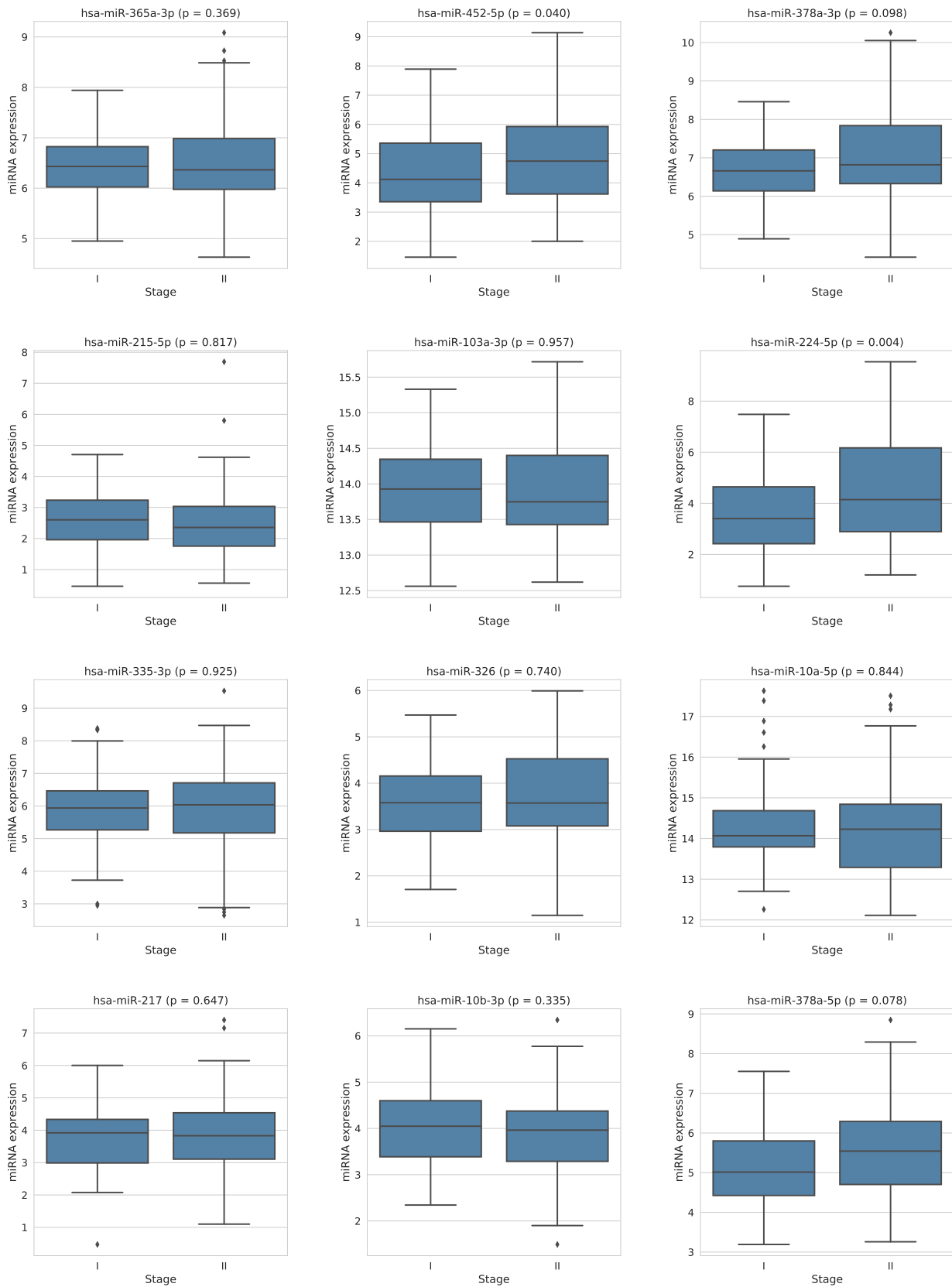


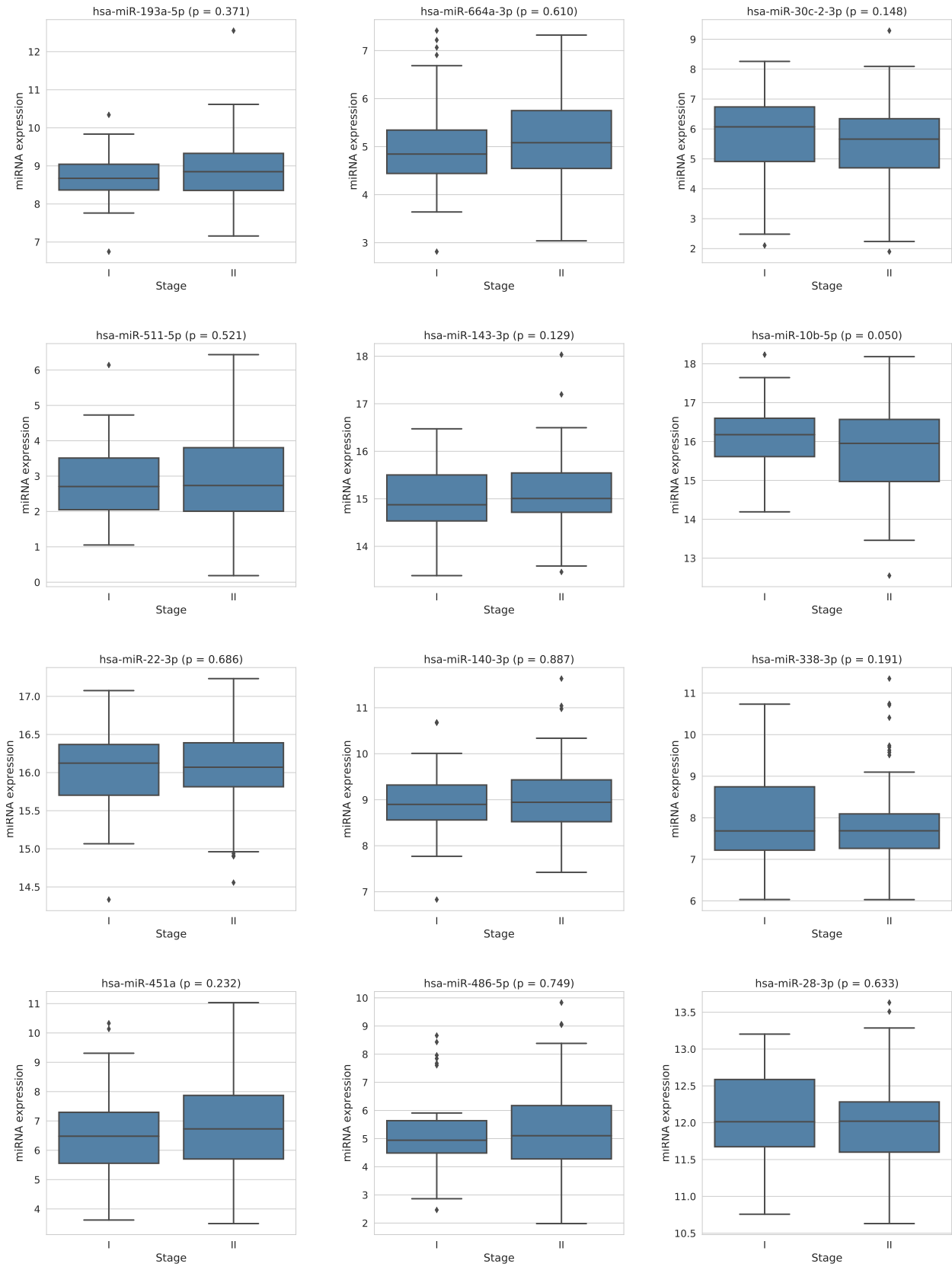
(b)

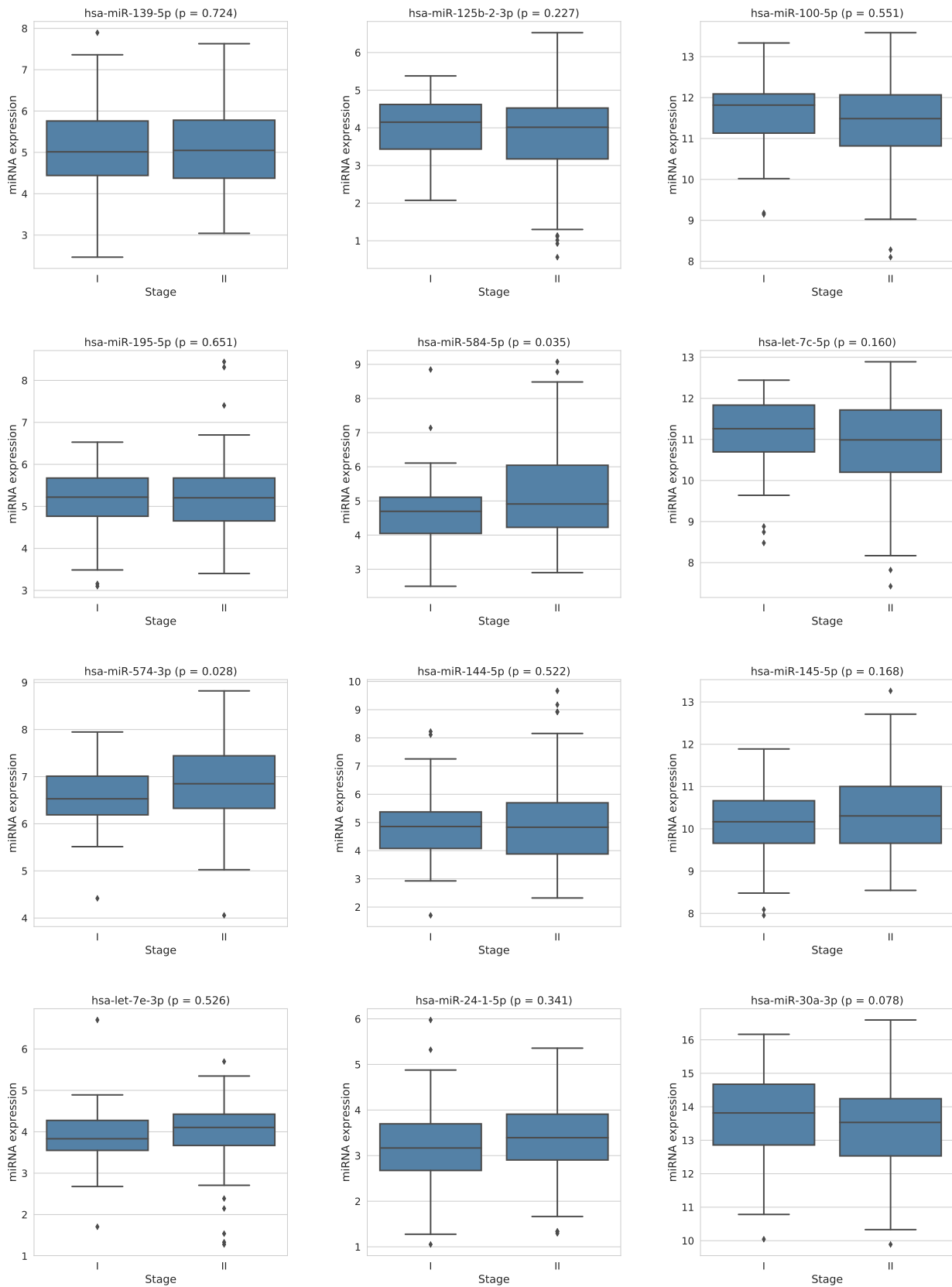


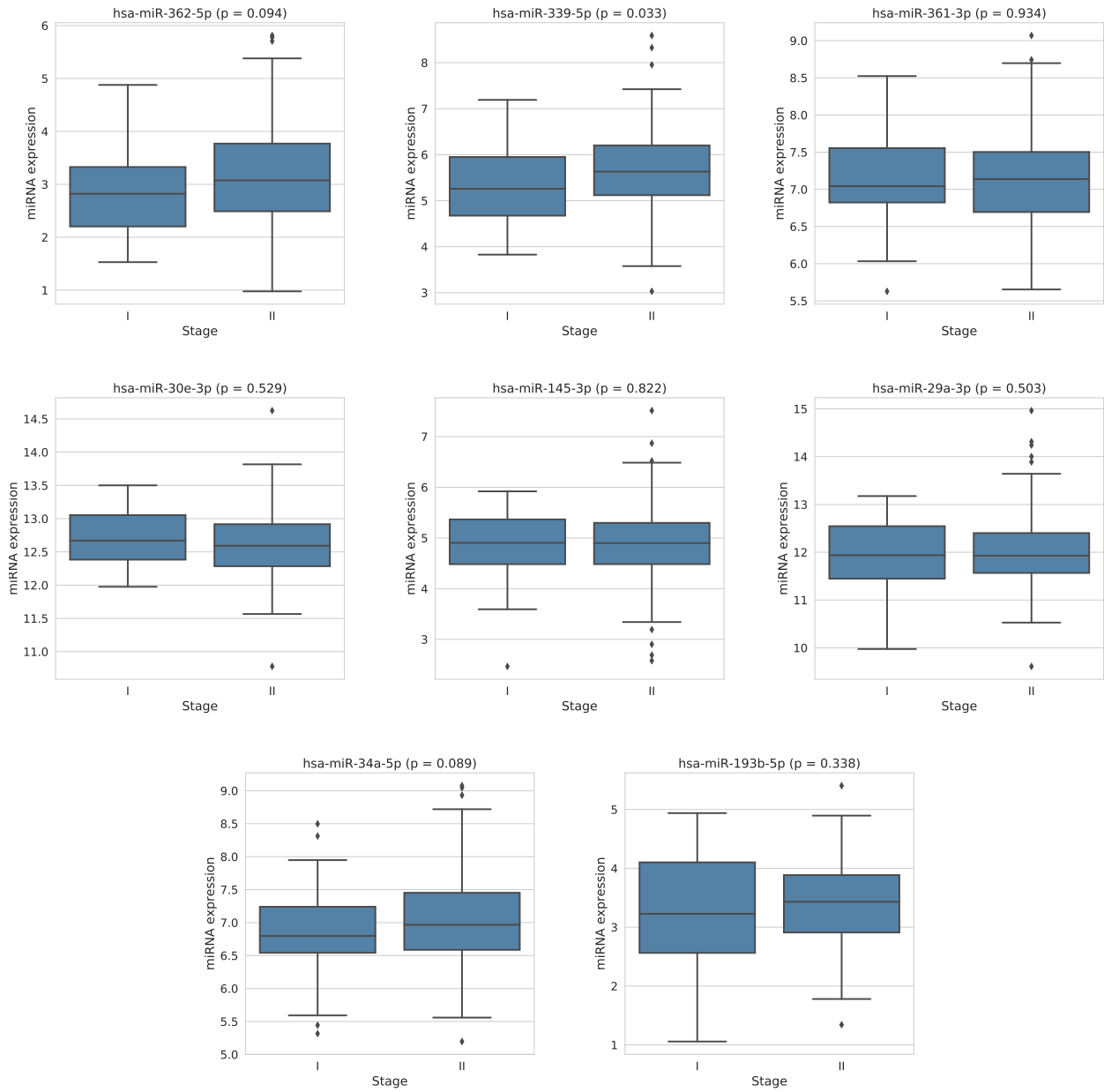
(d)

Figure S12: Clustering analysis of miRNA expression for (a) LA, (b) LB, (c) HER2-E and (d) BL subtypes with control patients. Color of the top bar represents group: red for cancer, green for control

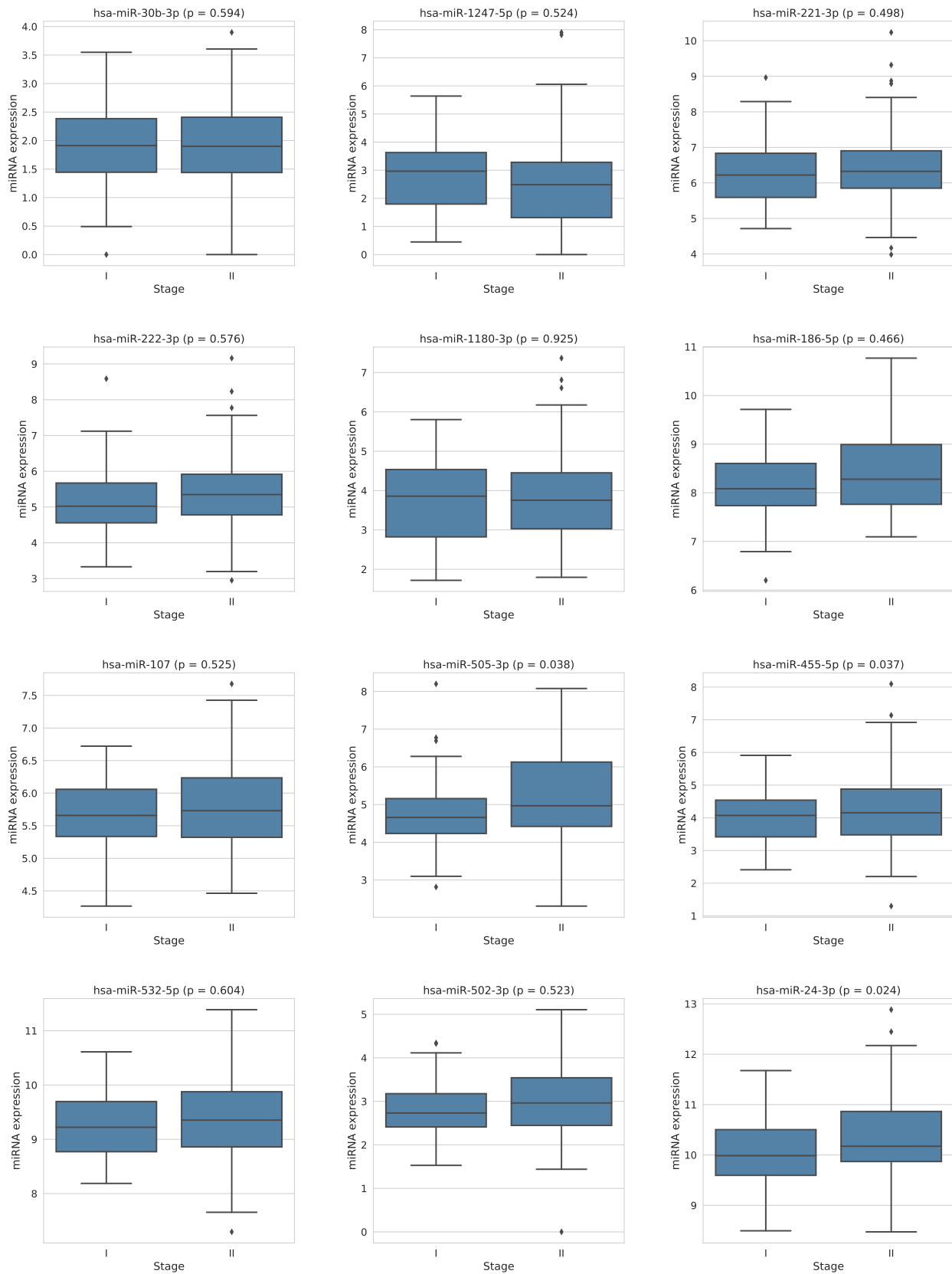




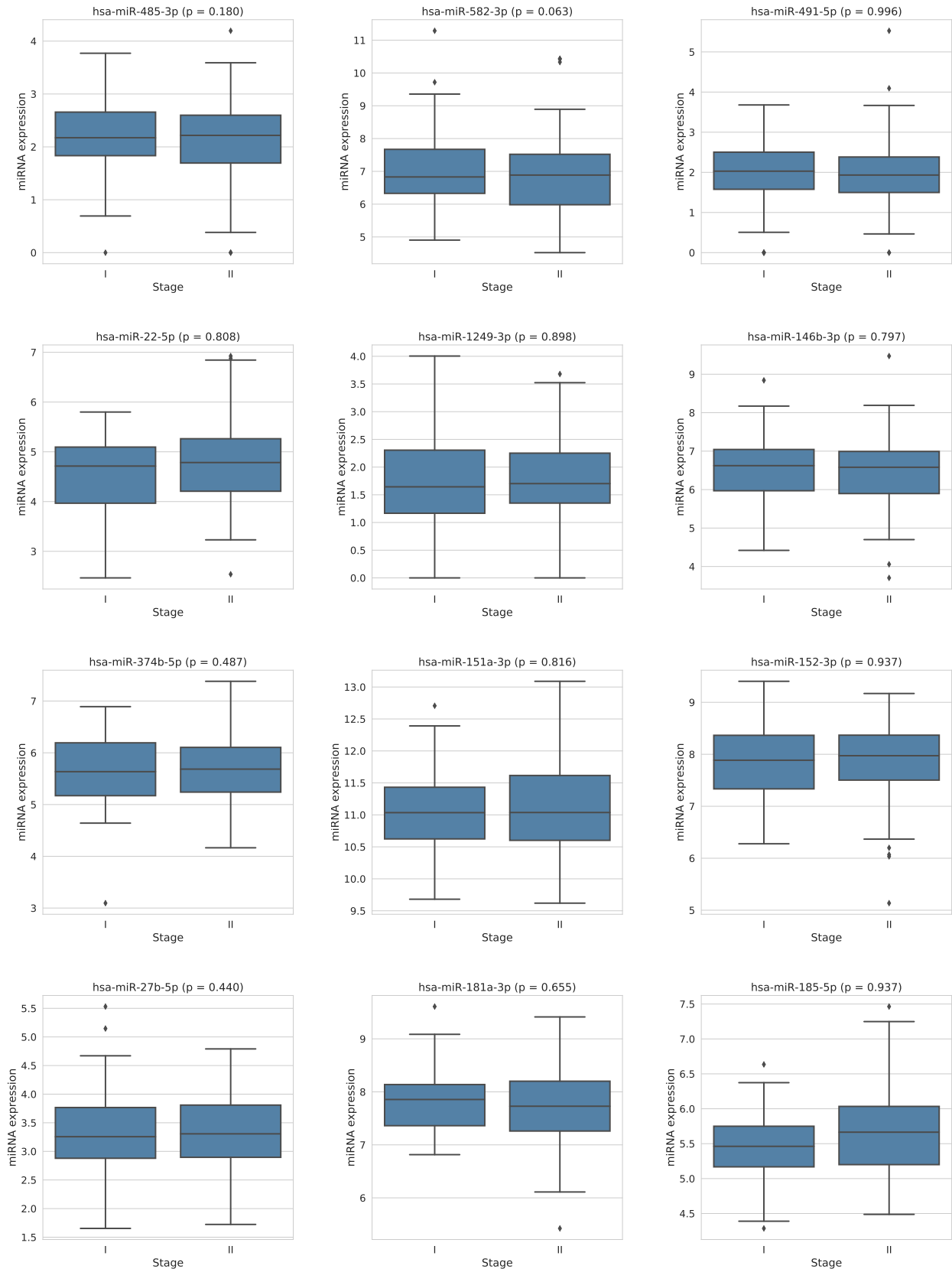




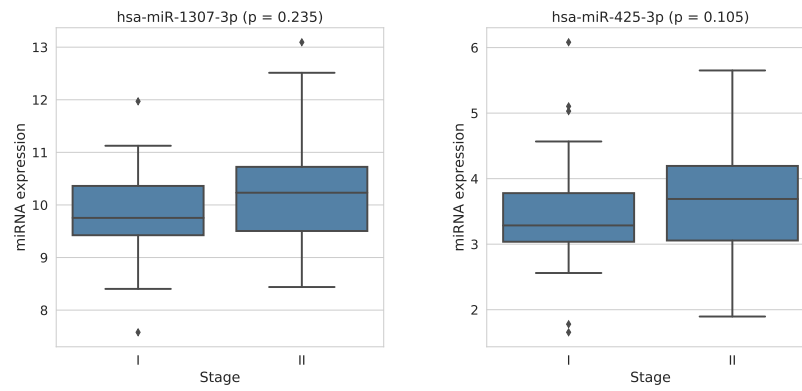
(a)



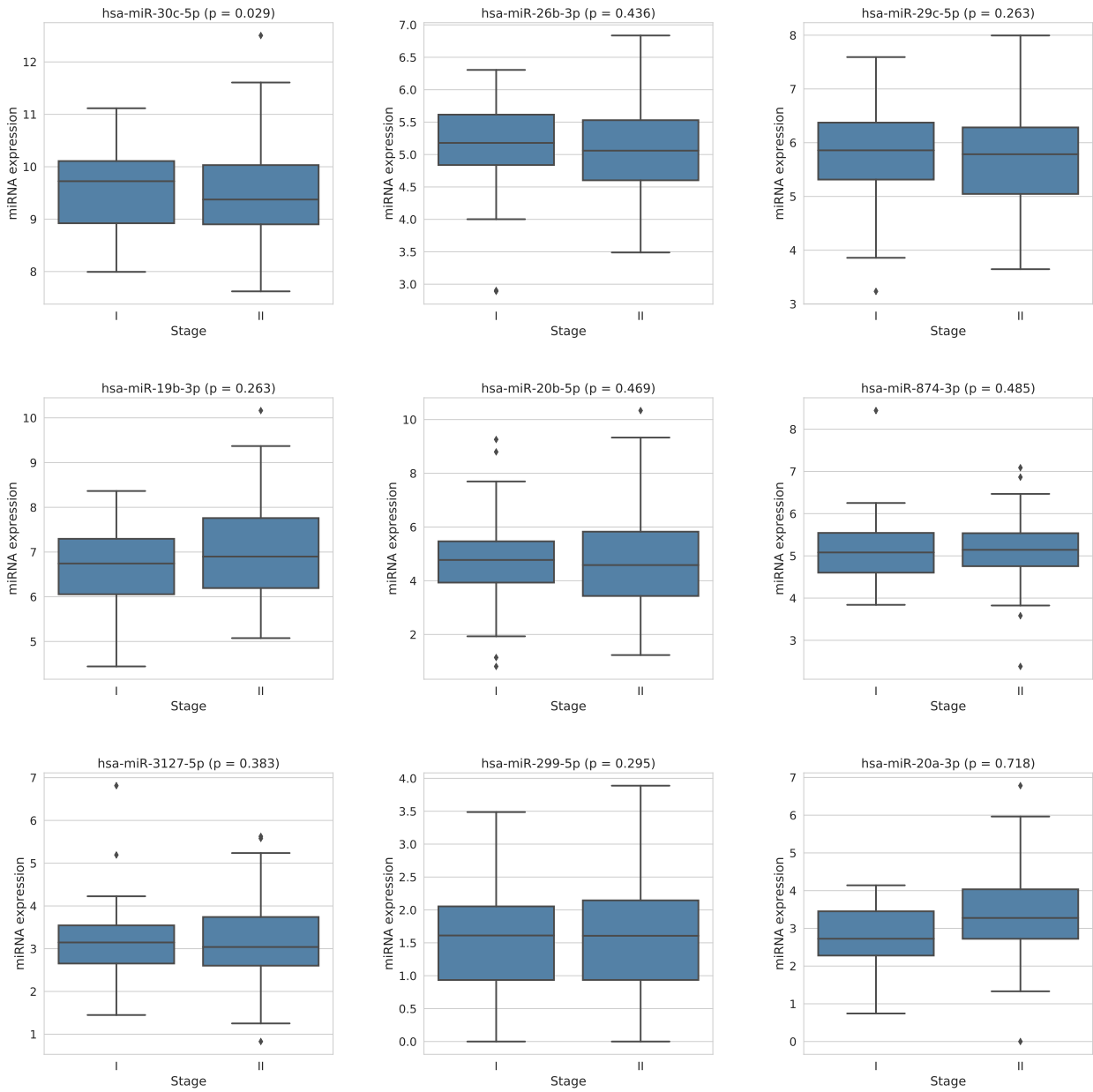
(b)



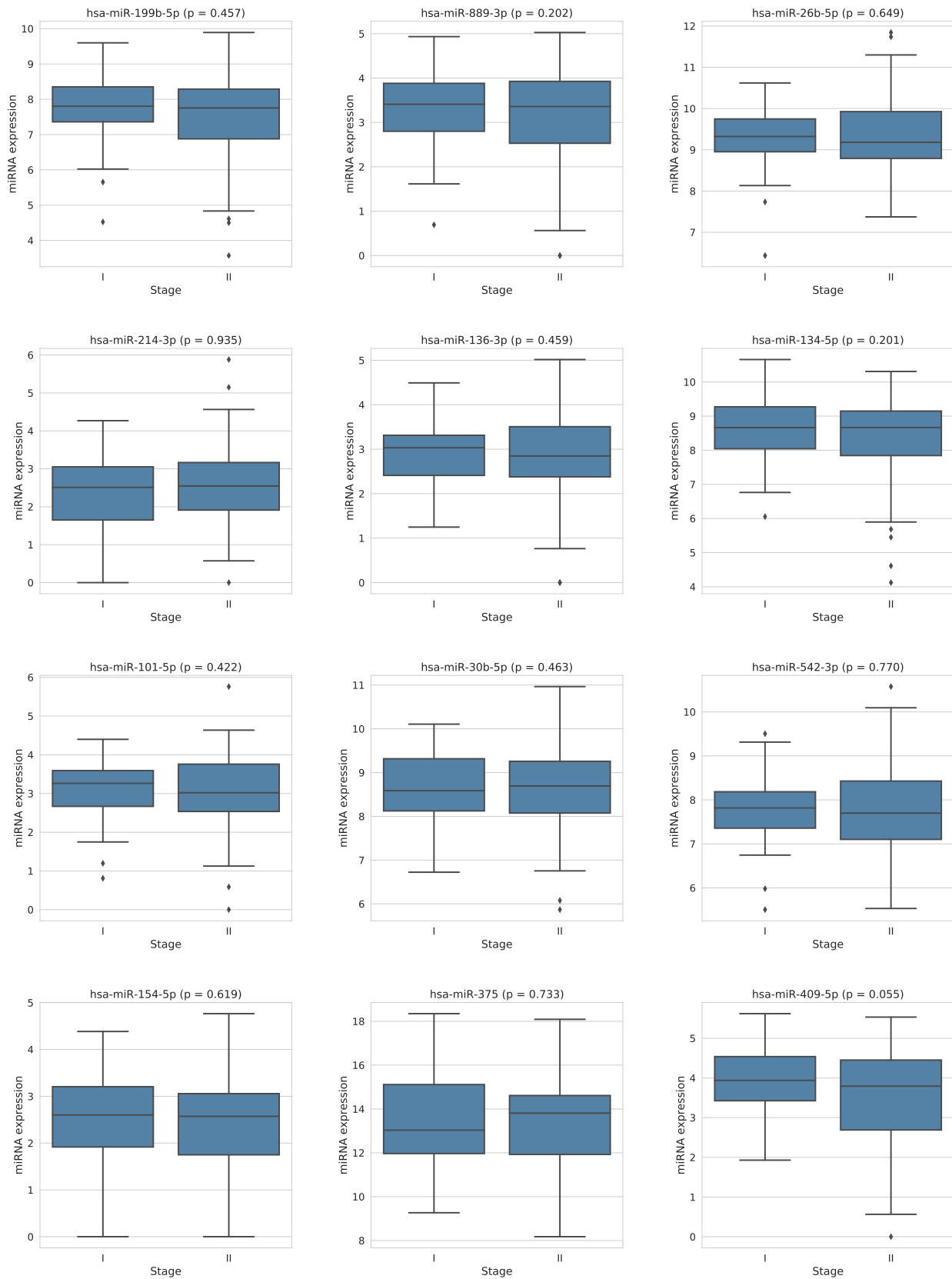


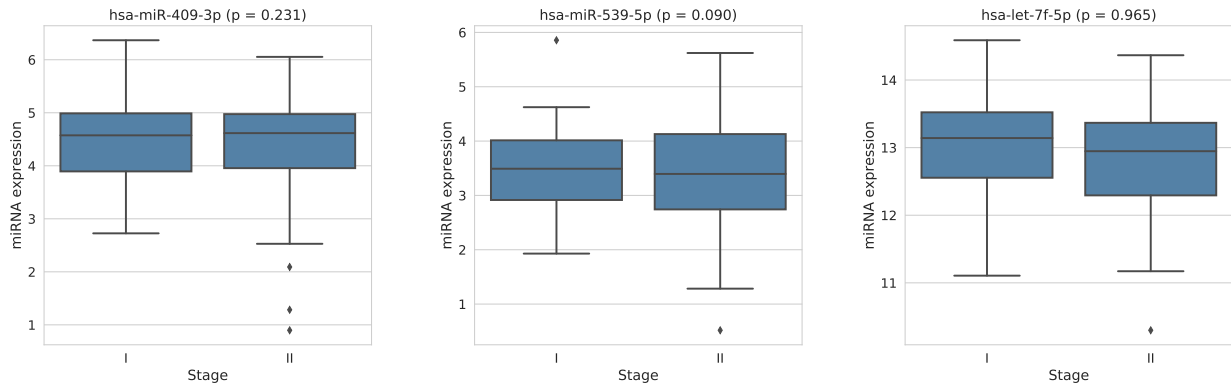


(c)



(d)

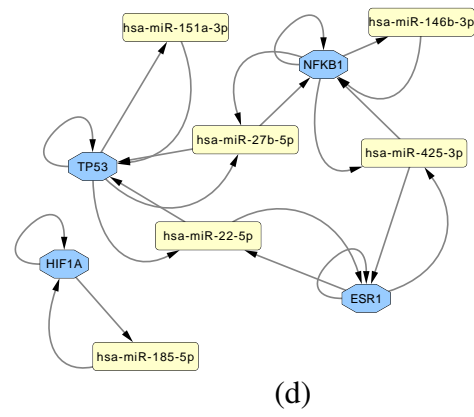
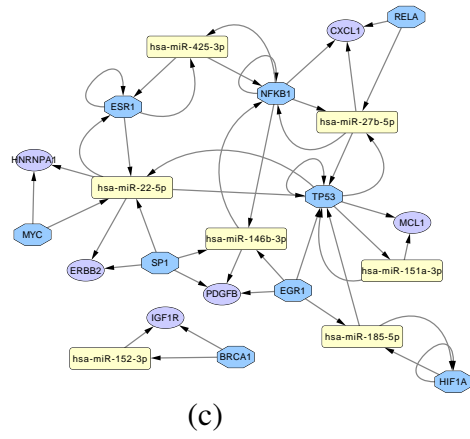
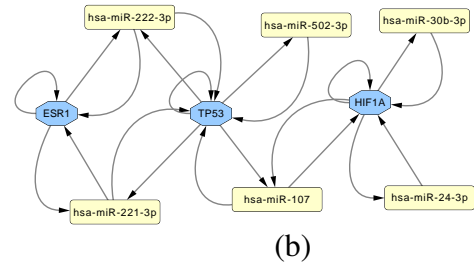
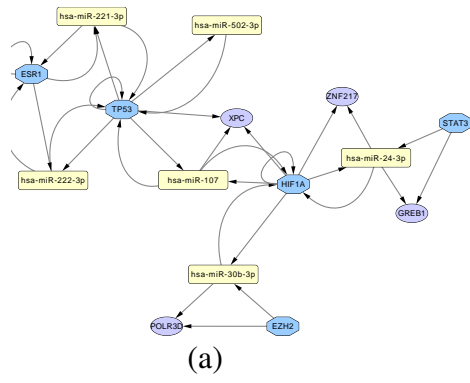




(e)

Figure S13: Boxplots showing the expression in cancer stages I and II for (a) 4-star miRNAs and 1-star miRNAs for (b) LA, (c) LB, (d) HER2-E and (e) BL





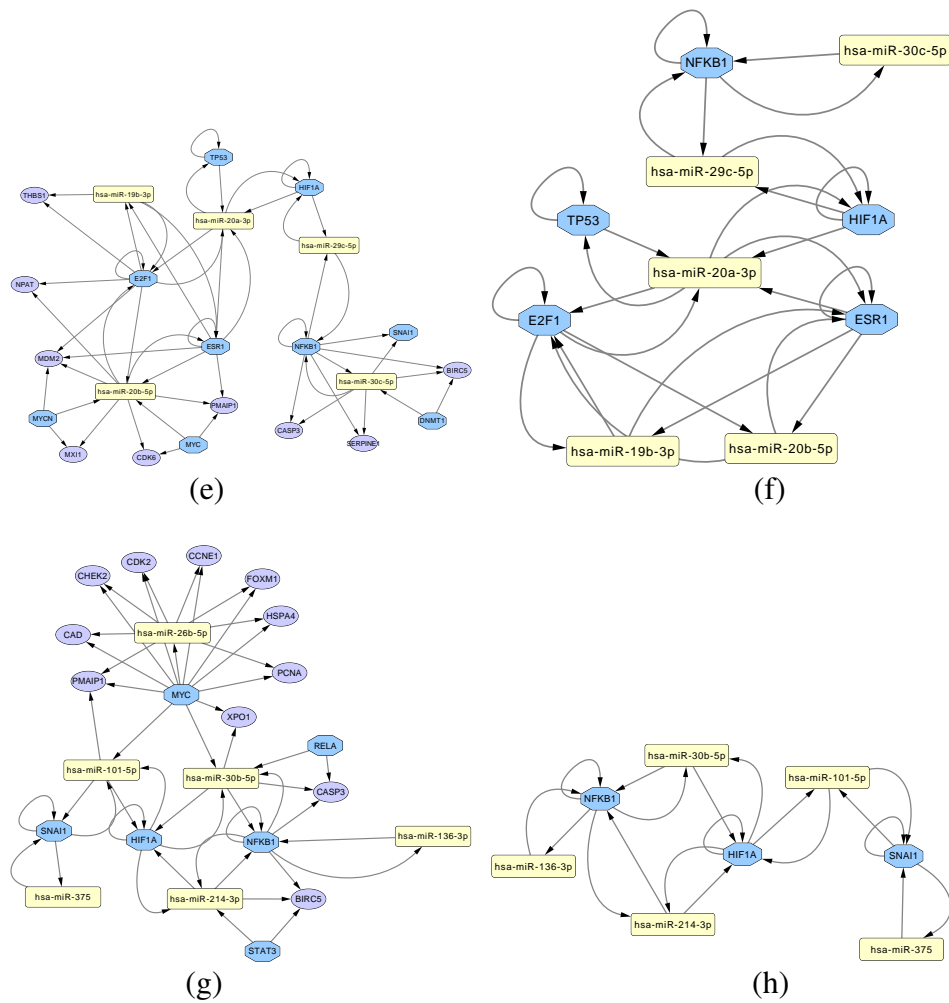
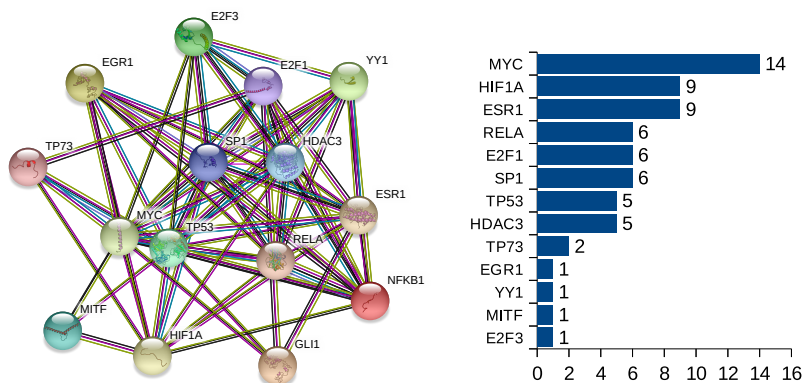
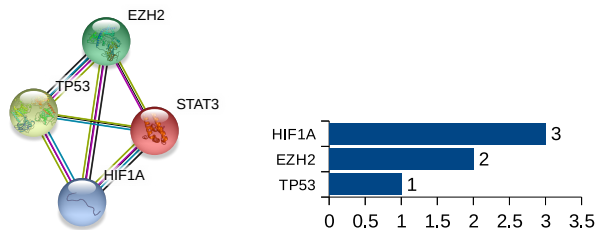


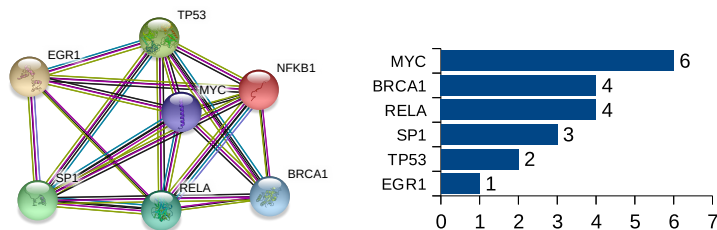
Figure S15: Regulatory networks of miRNAs (yellow rectangles), genes (purple ellipses) and transcription factors (blue octagons) for the 1-star miRNAs: LA (a, b), LB (c, d), HER2-E (e, f) and BL (g, h). Full networks are on the left and the ones only showing loops are on the right



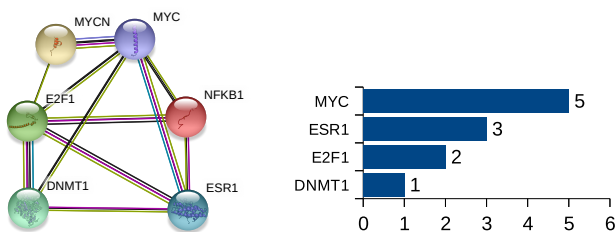
(a)



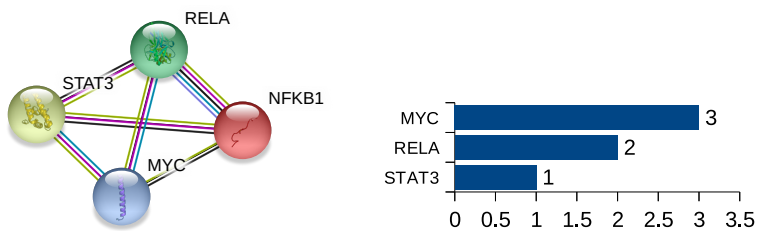
(b)



(c)



(d)



(e)

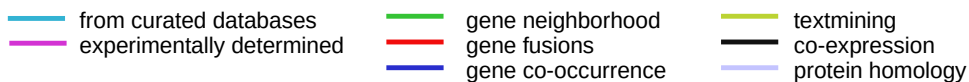


Figure S16: Protein-Protein Interaction networks for (a) 4-star and 1-star miRNAs for (b) LA, (c) LB, (d) HER2-E and (e) BL. The barplots show the degree (number of interactions) of each node in the Protein-Protein interaction networks. Colors of edges represent protein-protein association type, as indicated in the legend. Node colors are for aesthetic purposes