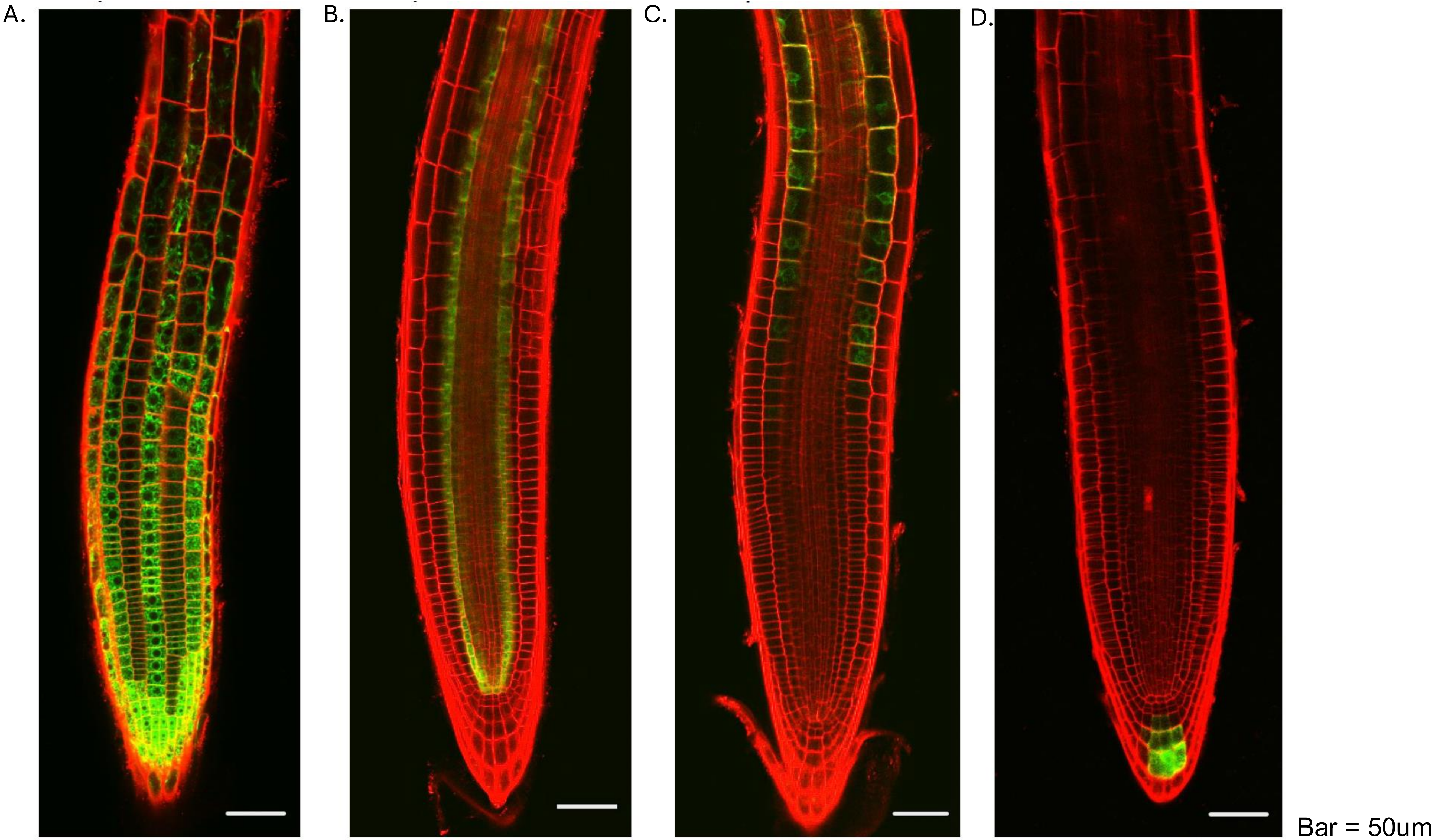


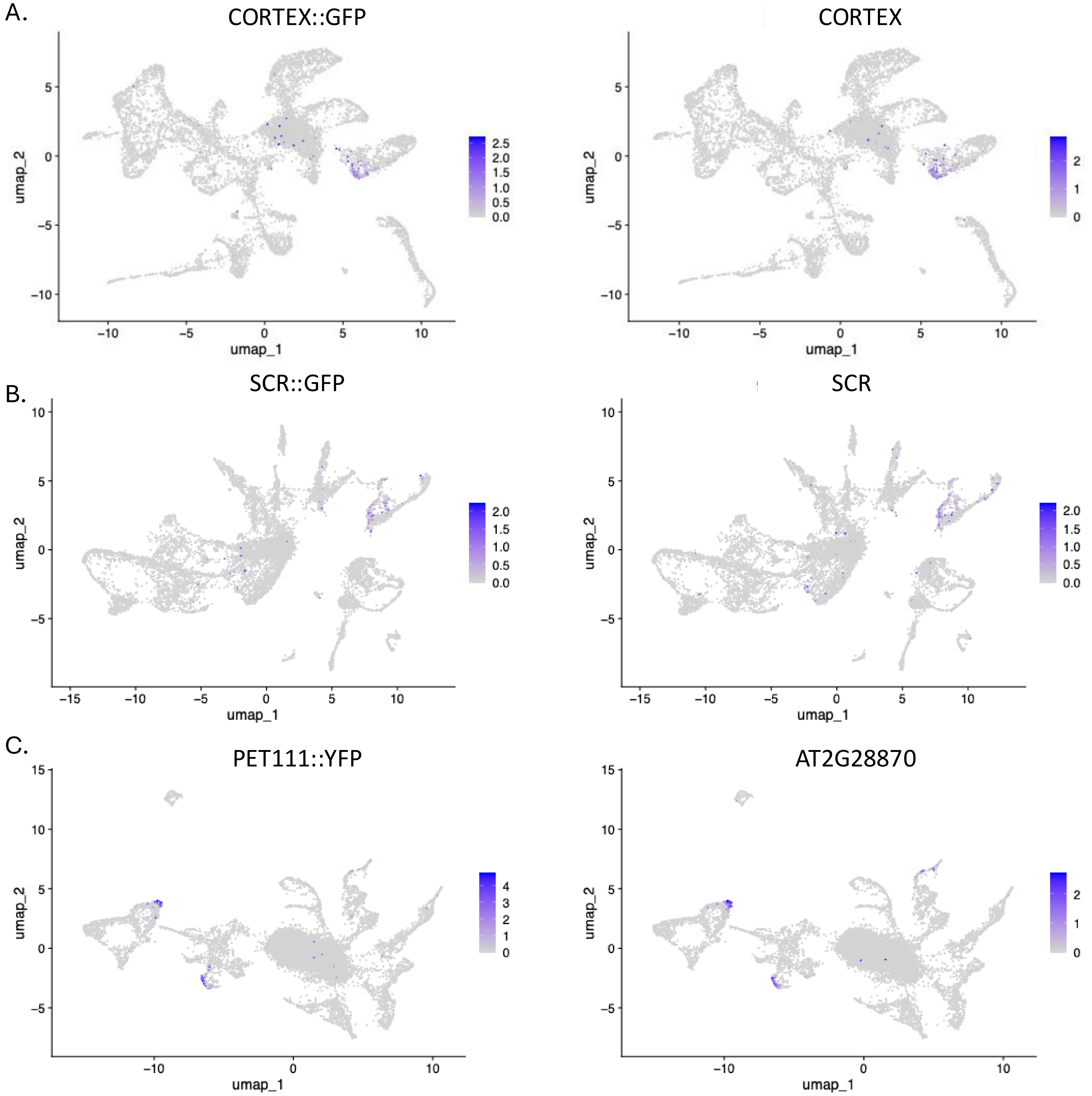
E.

Sample name	WER::GFP	SCR::GFP	CORTEX::GFP	PET111::YFP
Plant species	*Arabidopsis thaliana*	*Arabidopsis thaliana*	*Arabidopsis thaliana*	*Arabidopsis thaliana*
Experimental date	11/22/2022	11/22/2022	11/16/2022	11/16/2022
Estimated Number of Cells	10,017	11,098	10,992	7,720
Mean Reads per Cell	70,050	59,480	72,830	106,929
Median Genes per Cell	1,800	1,804	2,544	1,415
Fraction Reads in Cells	73.90%	64.00%	90.20%	81.00%
Number of Reads	701,692,194	660,110,948	800,543,009	825,493,937

Supplementary Figure 1.

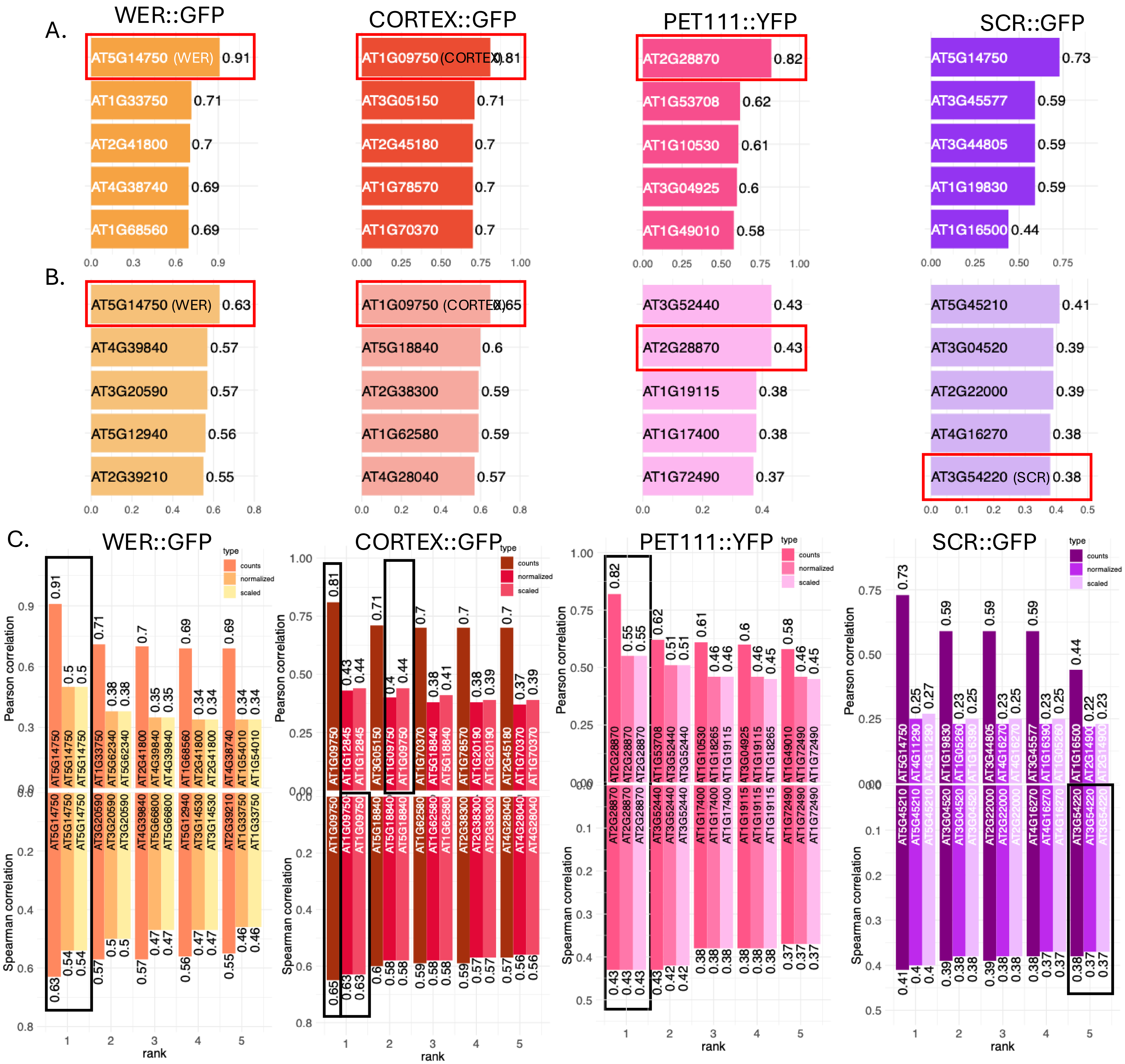
A-D. Confocal images showing WER::GFP fluorescence in the epidermis, SCR::GFP fluorescence in the endodermis, CORTEX::GFP fluorescence in the cortex, and PET111::YFP fluorescence near the root tip, respectively.

E. Total sequencing reads per sample.


A.
CORTEX::GFP
CORTEX
B.
SCR::GFP
SCR
C.
PET111::YFP
AT2G28870
umap_1
umap_2
2.5
2.0
1.5
1.0
0.5
0.0
4
3
2
1
0

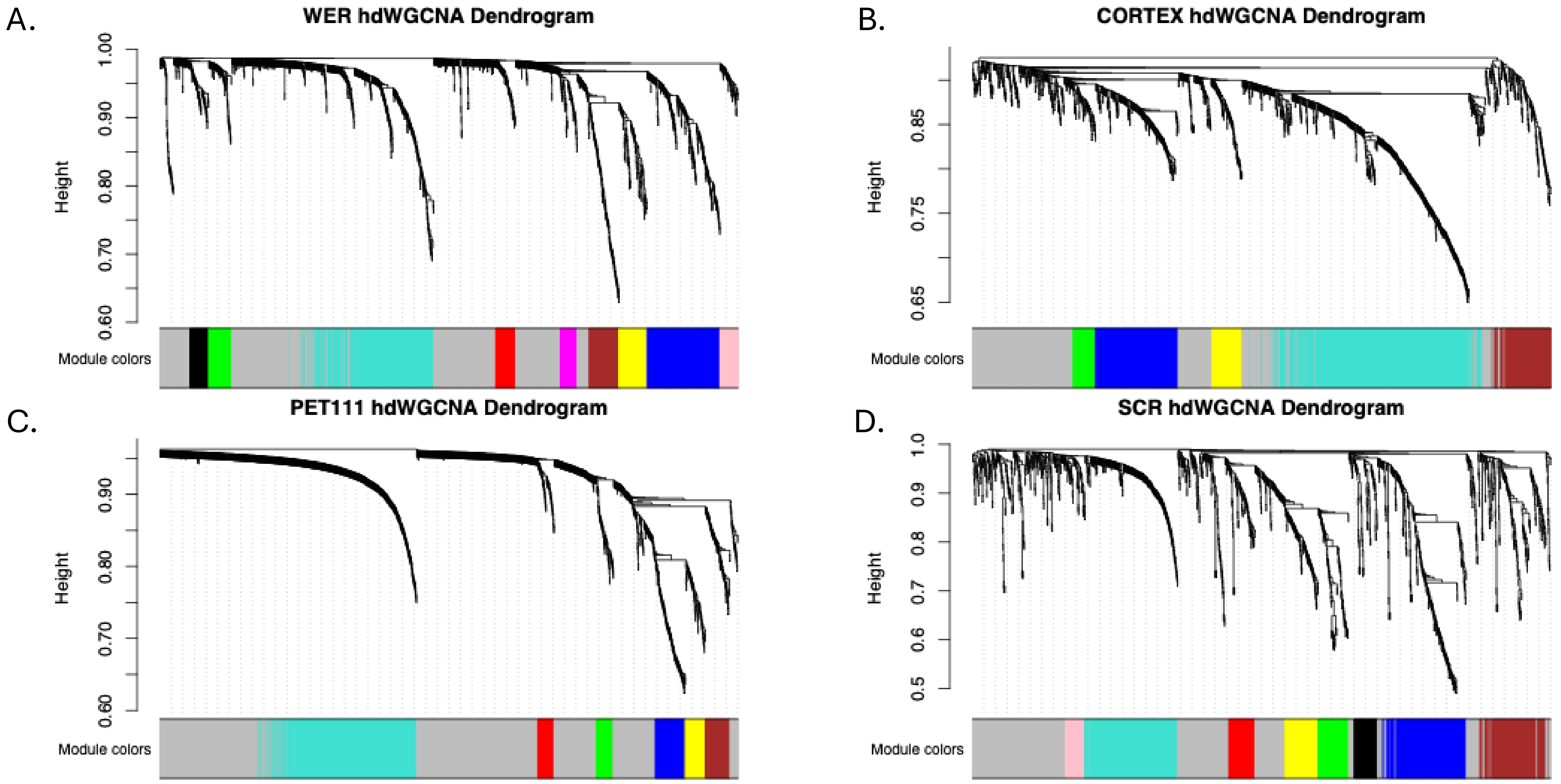
Supplementary Figure 2.

A. Feature plots displaying the gene expression patterns of CORTEX::GFP and the CORTEX gene, with each dot representing a cell. Darker purple indicates higher expression, while lighter purple indicates lower expression.
B. Feature plots showing the gene expression patterns of SCR::GFP and the SCR gene.
C. Feature plots illustrating the gene expression patterns of PET111::YFP and its corresponding gene, AT2G28870.


A.
WER::GFP
AT5G14750 (WER) 0.91
AT1G33750 0.71
AT2G41800 0.7
AT4G38740 0.69
AT1G68560 0.69
CORTEX::GFP
AT1G09750 (CORTEX) 0.81
AT3G05150 0.71
AT2G45180 0.7
AT1G78570 0.7
AT1G70370 0.7
PET111::YFP
AT2G28870 0.82
AT1G53708 0.62
AT1G10530 0.61
AT3G04925 0.6
AT1G49010 0.58
SCR::GFP
AT5G14750 0.73
AT3G45577 0.59
AT3G44805 0.59
AT1G19830 0.59
AT1G16500 0.44
B.
AT5G14750 (WER) 0.63
AT4G39840 0.57
AT3G20590 0.57
AT5G12940 0.56
AT2G39210 0.55
AT1G09750 (CORTEX) 0.65
AT5G18840 0.6
AT2G38300 0.59
AT1G62580 0.59
AT4G28040 0.57
AT3G52440 0.43
AT2G28870 0.43
AT1G19115 0.38
AT1G17400 0.38
AT1G72490 0.37
AT5G45210 0.41
AT3G04520 0.39
AT2G22000 0.39
AT4G16270 0.38
AT3G54220 (SCR) 0.38
C.
WER::GFP
CORTEX::GFP
PET111::YFP
SCR::GFP
Pearson correlation
Spearman correlation
rank
type
counts
normalized
scaled

Supplementary Figure 3.

A. Bar charts showing the top 5 genes and their correlation to the promoter-reporter genes in four samples using the Pearson method. The red box highlights the native or corresponding genes. While WER::GFP, CO2::GFP, and PET111::YFP exhibit the highest correlation with their native genes, SCR::GFP shows a weaker correlation, ranking 163rd with its native gene.
B. Bar charts displaying the top 5 genes and their correlation to the promoter-reporter genes in four samples using the Spearman method. The red box highlights the native or corresponding genes.
C. Bar charts showing the top 5 genes with the highest correlation to their promoter-reporter genes across three data types (counts, normalized, and scaled) in four samples. The black box highlights the rank of their native genes in each sample. The top bar chart is based on Pearson correlation, while the bottom bar chart uses Spearman correlation.



E.

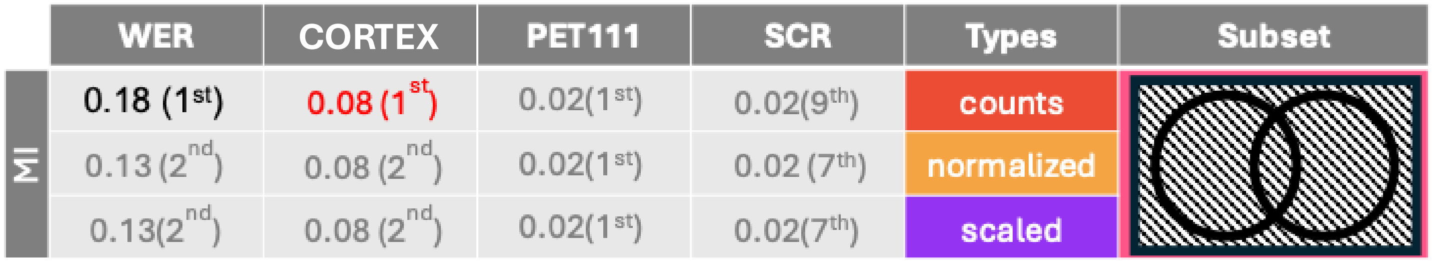
Sample	Gene	Module	TOM value	Rank
WER	WER	pink	0.048	2
WER	WER::GFP	pink		
CORTEX	CORTEX	brown	0.192	15
CORTEX	CORTEX::GFP	brown		
PET111	AT2G28870	red	0.106	8
PET111	PET111::YFP	red		
SCR	SCR	brown	0.029	118
SCR	SCR::GFP	brown		

Supplementary Figure 4.

A-D. Dendrograms showing gene modules of the top 2,000 highly variable genes in four samples: WER, CORTEX, PET111, and SCR, respectively.

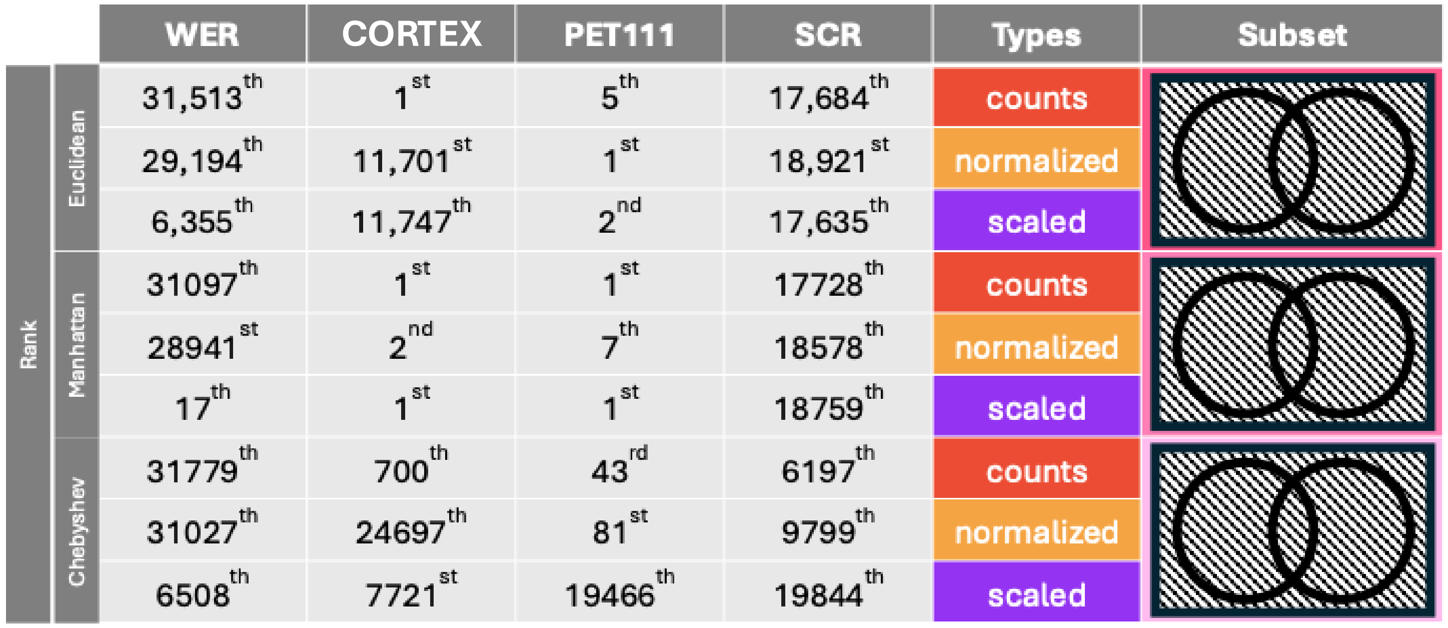
E. Table showing the assigned module, TOM value, and rank between promoter-reporter and corresponding native genes in all four samples.

A.



	WER	CORTEX	PET111	SCR	Types	Subset
MI	0.18 (1st)	0.08 (1st)	0.02(1st)	0.02(9th)	counts	
	0.13 (2nd)	0.08 (2nd)	0.02(1st)	0.02 (7th)	normalized	
	0.13(2nd)	0.08 (2nd)	0.02(1st)	0.02(7th)	scaled	

B.



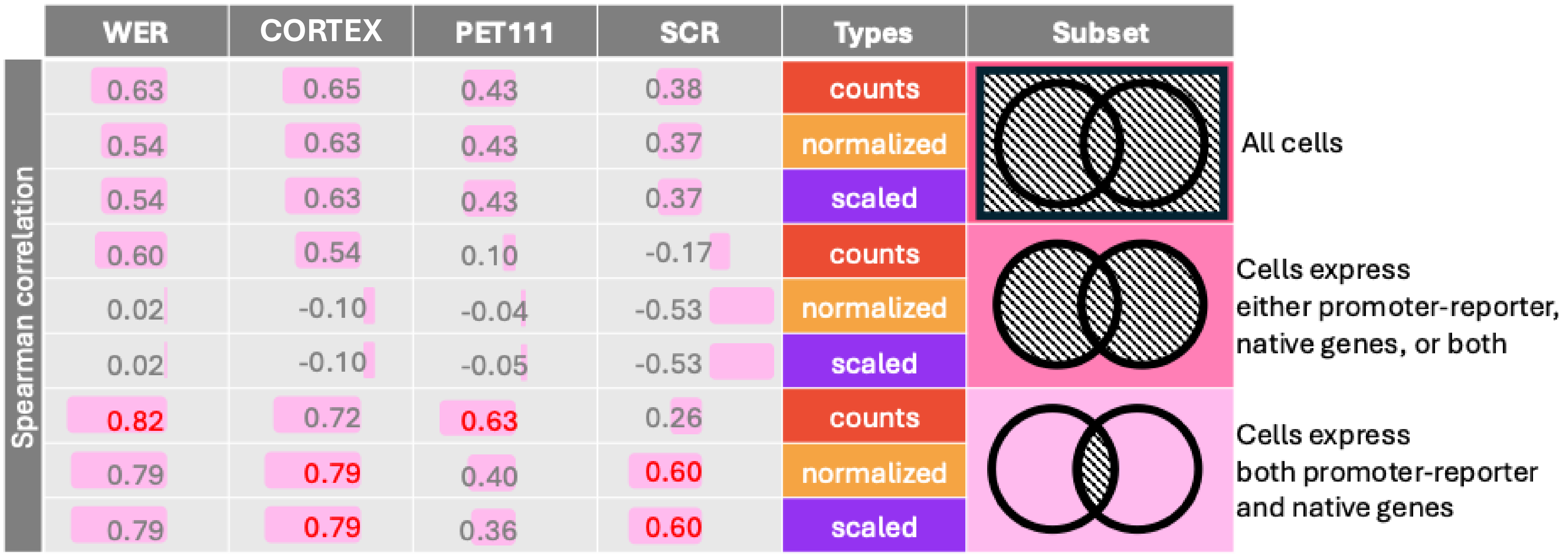
		WER	CORTEX	PET111	SCR	Types	Subset
Rank	Euclidean	31,513th	1st	5th	17,684th	counts	
		29,194th	11,701st	1st	18,921st	normalized	
		6,355th	11,747th	2nd	17,635th	scaled	
	Manhattan	31097th	1st	1st	17728th	counts	
		28941st	2nd	7th	18578th	normalized	
		17th	1st	1st	18759th	scaled	
	Chebyshev	31779th	700th	43rd	6197th	counts	
		31027th	24697th	81st	9799th	normalized	
		6508th	7721st	19466th	19844th	scaled	

Supplementary Figure 5.

A. This table presents the mutual information between promoter-reporter genes and their native counterparts across various samples, along with the rank of this similarity compared to all other genes in the genome. The analysis was performed using three data types: counts, normalized, and scaled.

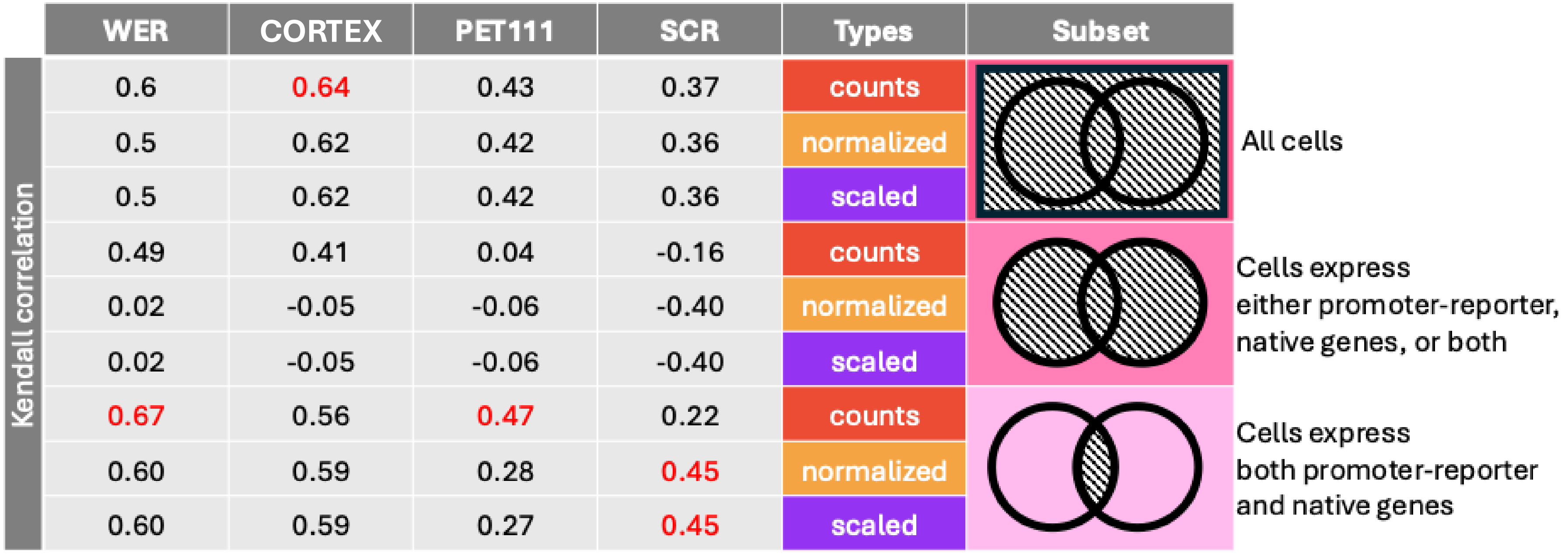
B. The table presents the rank of native genes in relation to their corresponding promoter-reporter genes using three distance metrics: Euclidean, Manhattan, and Chebyshev, measured across the data types counts, normalized, and scaled.

A.



	WER	CORTEX	PET111	SCR	Types	Subset
Spearman correlation	0.63	0.65	0.43	0.38	counts	All cells
	0.54	0.63	0.43	0.37	normalized	
	0.54	0.63	0.43	0.37	scaled	
	0.60	0.54	0.10	-0.17	counts	Cells express either promoter-reporter, native genes, or both
	0.02	-0.10	-0.04	-0.53	normalized	
	0.02	-0.10	-0.05	-0.53	scaled	
	0.82	0.72	0.63	0.26	counts	Cells express both promoter-reporter and native genes
	0.79	0.79	0.40	0.60	normalized	
	0.79	0.79	0.36	0.60	scaled	

B.

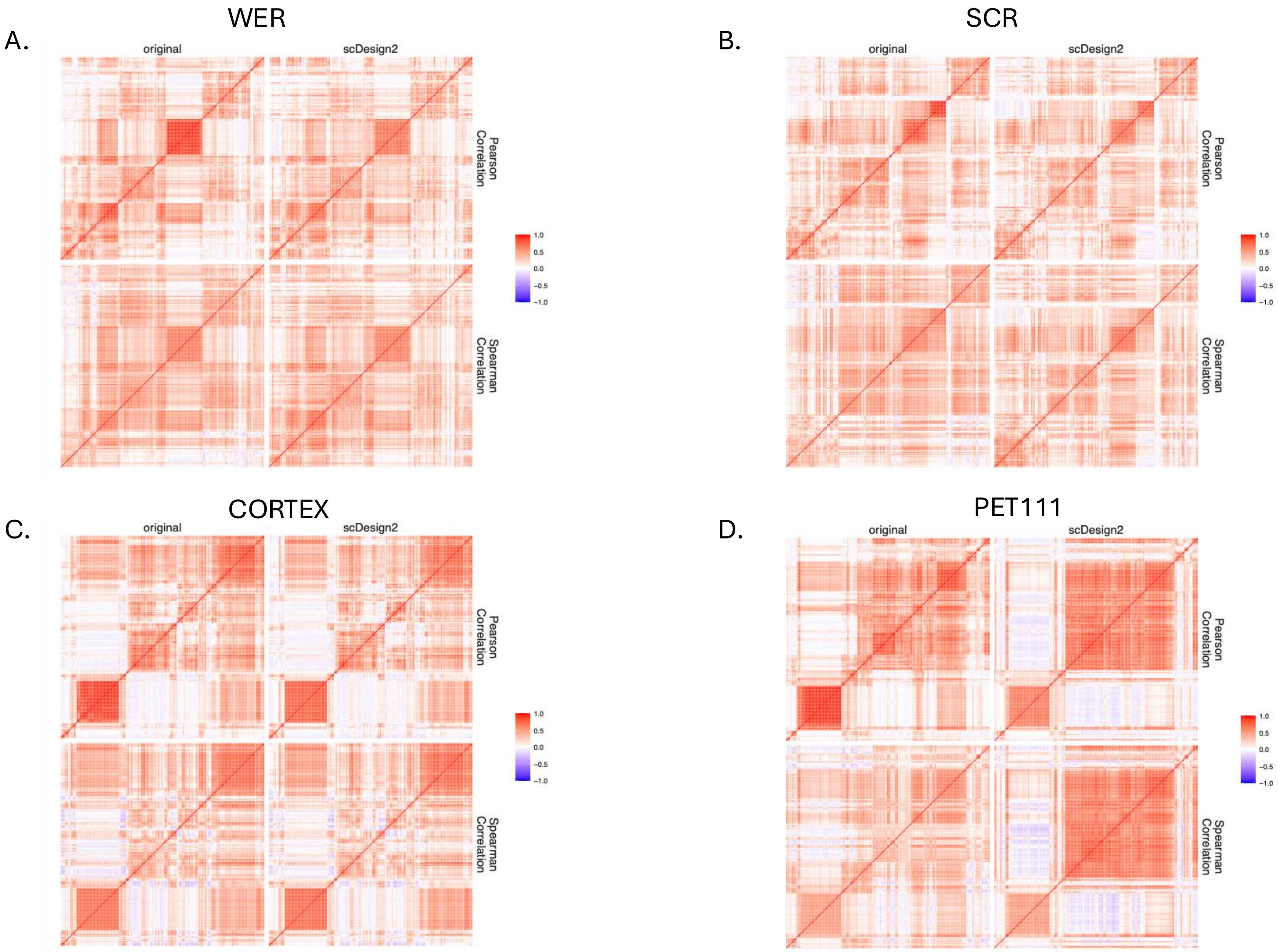


	WER	CORTEX	PET111	SCR	Types	Subset
Kendall correlation	0.6	0.64	0.43	0.37	counts	All cells
	0.5	0.62	0.42	0.36	normalized	
	0.5	0.62	0.42	0.36	scaled	
	0.49	0.41	0.04	-0.16	counts	Cells express either promoter-reporter, native genes, or both
	0.02	-0.05	-0.06	-0.40	normalized	
	0.02	-0.05	-0.06	-0.40	scaled	
	0.67	0.56	0.47	0.22	counts	Cells express both promoter-reporter and native genes
	0.60	0.59	0.28	0.45	normalized	
	0.60	0.59	0.27	0.45	scaled	

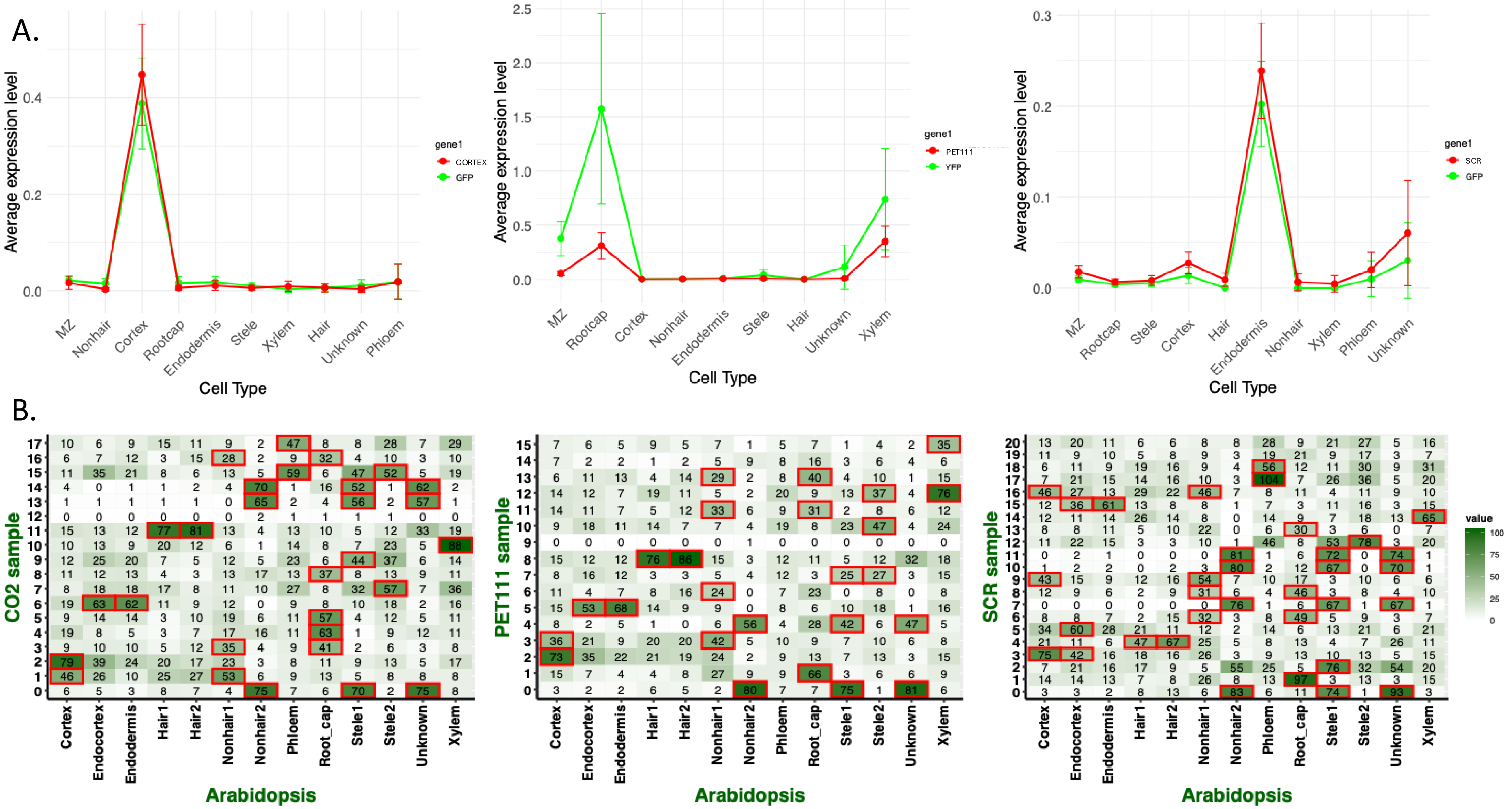
Supplementary Figure 6.

A. The table shows the correlation between promoter-reporter genes and their native counterparts across different data types (counts, normalized, and scaled), various subsets (all cells, cells expressing either promoter-reporter, native genes, or both, and cells expressing both promoter-reporter and native genes), and different samples. Red text highlights the dataset with the highest value in each sample. Correlation values in this table were measured using the Spearman method.

B. The table shows the correlation between promoter-reporter genes and their native counterparts across different data types (counts, normalized, and scaled), various subsets (all cells, cells expressing either promoter-reporter, native genes, or both, and cells expressing both promoter-reporter and native genes), and different samples. Red text highlights the dataset with the highest value in each sample. Correlation values in this table were measured using the Kendall method.



Supplementary Figure 7.
A–D: Gene-gene correlation analysis of the top 200 highly expressed genes in WER (A), SCR (B), CORTEX (C), and PET111 (D) samples. Correlation coefficients were computed using both Pearson and Spearman methods in the original single-cell RNA-seq data and the scDesign2 simulated data.



CO2 sample	Cortex	Endocortex	Endodermis	Hair1	Hair2	Nonhair1	Nonhair2	Phloem	Root_cap	Stele1	Stele2	Unknown	Xylem
17	10	6	9	15	11	9	2	47	8	8	28	7	29
16	6	7	12	3	15	28	2	9	32	4	10	3	10
15	11	35	21	8	6	13	5	59	6	47	52	5	19
14	4	0	1	1	2	4	70	1	16	52	1	62	2
13	1	1	1	1	1	0	65	2	4	56	2	57	1
12	0	0	0	0	0	0	2	1	1	1	1	0	0
11	15	13	12	77	81	13	4	13	10	5	12	33	19
10	10	13	9	20	12	6	1	14	8	7	23	5	88
9	12	25	20	7	5	12	5	23	6	44	37	6	14
8	11	12	13	4	3	13	17	13	37	8	13	9	11
7	8	18	18	17	8	11	10	27	8	32	57	7	36
6	19	63	62	11	9	12	0	9	8	10	18	2	16
5	9	14	14	3	10	19	6	11	57	4	12	5	11
4	19	8	5	3	7	17	16	11	63	1	9	12	11
3	9	10	10	5	12	35	4	9	41	2	6	3	8
2	79	39	24	20	17	23	3	8	11	9	13	5	17
1	46	26	10	25	27	53	6	9	13	5	8	8	8
0	6	5	3	8	7	4	75	7	6	70	2	75	8

PET111 sample	Cortex	Endocortex	Endodermis	Hair1	Hair2	Nonhair1	Nonhair2	Phloem	Root_cap	Stele1	Stele2	Unknown	Xylem
15	7	6	5	9	5	7	1	5	7	1	4	2	35
14	7	2	2	1	2	5	9	8	16	3	6	4	6
13	6	11	13	4	14	29	2	8	40	4	10	1	15
12	14	12	7	19	11	5	2	20	9	13	37	4	76
11	14	11	7	5	12	33	6	9	31	2	8	6	12
10	9	18	11	14	7	7	4	19	8	23	47	4	24
9	0	0	0	0	0	0	1	0	0	0	0	0	0
8	15	12	12	76	86	15	3	12	11	5	12	32	18
7	8	16	12	3	3	5	4	12	3	25	27	3	15
6	11	4	7	8	16	24	0	7	23	0	8	0	8
5	15	53	68	14	9	9	0	8	6	10	18	1	16
4	8	2	5	1	0	6	56	4	28	42	6	47	5
3	36	21	9	20	20	42	5	10	9	7	10	6	8
2	73	35	22	21	19	24	2	9	13	7	13	3	15
1	15	7	4	4	8	27	9	9	66	3	6	6	9
0	3	2	2	6	5	2	80	7	7	75	1	81	6

SCR sample	Cortex	Endocortex	Endodermis	Hair1	Hair2	Nonhair1	Nonhair2	Phloem	Root_cap	Stele1	Stele2	Unknown	Xylem
20	13	20	11	6	6	8	8	28	9	21	27	5	16
19	11	9	10	5	8	9	3	19	21	9	17	3	7
18	6	11	9	19	16	9	4	56	12	11	30	9	31
17	7	21	15	14	16	10	3	104	7	26	36	5	20
16	46	27	13	29	22	46	7	8	11	4	11	9	10
15	12	36	61	13	8	8	1	7	3	11	16	3	15
14	12	11	14	26	14	8	0	14	9	7	18	13	65
13	8	8	11	5	10	22	0	6	30	3	6	0	7
12	11	22	15	3	3	10	1	46	8	53	78	3	20
11	0	2	1	0	0	0	81	1	6	72	0	74	1
10	1	2	2	0	2	3	80	2	12	67	0	70	1
9	43	15	9	12	16	54	7	10	17	3	10	6	6
8	12	9	6	2	9	31	6	4	46	3	8	4	10
7	0	0	0	0	0	0	76	1	6	67	1	67	1
6	8	8	7	2	15	32	3	8	49	5	9	3	7
5	34	60	28	21	11	12	2	14	6	13	21	6	21
4	21	11	6	47	67	25	5	8	13	4	7	26	11
3	75	42	16	18	16	26	3	9	13	10	13	5	15
2	7	21	16	17	9	5	55	25	5	76	32	54	20
1	14	14	13	7	8	26	8	13	97	3	13	3	15
0	3	3	2	8	13	6	83	6	11	74	1	93	3

Supplementary Figure 8.

A. The line plots display the average gene expression of promoter-reporter genes and their native counterparts across all cell-type clusters in each sample.

B. The heatmaps present the OMG cell type predictions for each cluster in each sample, with the red box highlighting the predicted cell type, tested by Fisher's exact test.

	WER::GFP, default cluster					WER::GFP, cell type				
	Pearson	Spearman	Kendall	P-sctransform	CSCORE	Pearson	Spearman	Kendall	P-sctransform	CSCORE
Count data	0.998 4.93761e-19	0.96 1.35929e-06	0.85 2.394031e-08	0.992 7.039501e-15	1 NA	0.993 8.200954e-08	0.95 0.0003527337	0.83 0.0008542769	0.99 9.23807e-07	1 NA
Normalized	0.98 1.337122e-12	0.93 0	0.82 1.257166e-07			0.97 1.903523e-05	0.82 0.0107694	0.72 0.005886243		
Scaled	0.98 1.337122e-12	0.93 0	0.82 1.257166e-07			0.97 1.903523e-05	0.82 0.0107694	0.72 0.005886243		

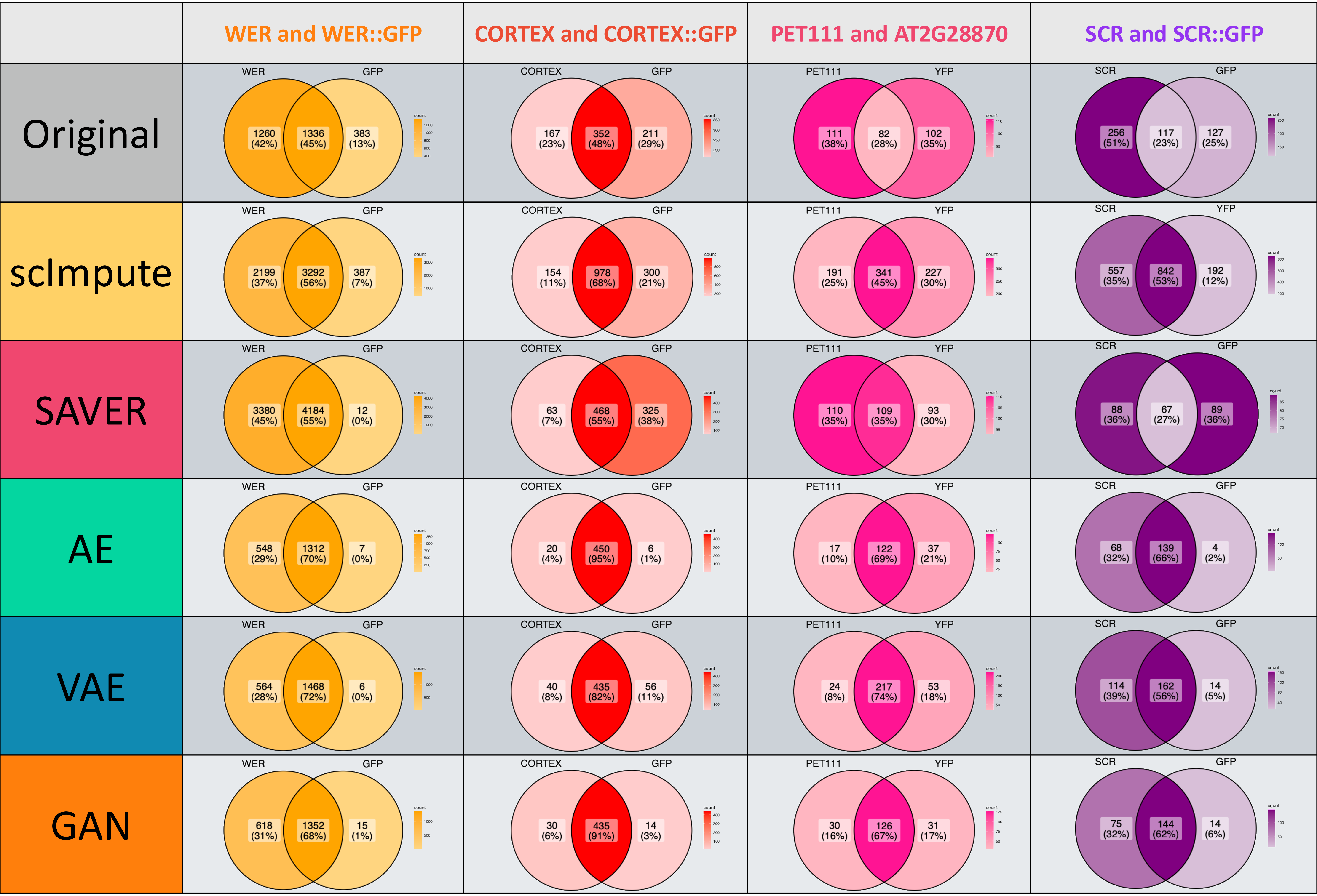
	SCR::GFP, default cluster					SCR::GFP, cell type				
	Pearson	Spearman	Kendall	P-sctransform	CSCORE	Pearson	Spearman	Kendall	P-sctransform	CSCORE
Count data	0.94 3.676867e-10	0.63 0.002336178	0.5 0.00226473	0.8 1.260116e-05	1 NA	0.995 1.808614e-09	0.93 0.0001170712	0.87 0.0005610084	0.9 0.0002845613	1 NA
Normalized	0.83 2.920657e-06	0.66 0.001020686	0.55 0.000775149			0.94 4.225192e-05	0.9 0.0003592905	0.83 0.001082229		
Scaled	0.84 1.946442e-06	0.68 0.0006462133	0.56 0.0006188882			0.96 1.560765e-05	0.9 0.0003592905	0.83 0.001082229		

	CORTEX::GFP, default cluster					CORTEX::GFP, cell type				
	Pearson	Spearman	Kendall	P-sctransform	CSCORE	Pearson	Spearman	Kendall	P-sctransform	CSCORE
Count data	0.99 9.831794e-17	0.3 0.2324642	0.3 0.08122441	0.99 8.103168e-14	1 NA	0.999 3.627519e-12	0.64 0.05444507	0.42 0.1083135	0.995 2.525452e-09	1 NA
Normalized	0.98 6.964249e-12	0.43 0.0776131	0.31 0.06884504			0.99 2.624119e-08	0.54 0.1132981	0.38 0.1557418		
Scaled	0.98 6.558317e-12	0.41 0.0876051	0.18 0.2885367			0.99 2.02694e-08	0.54 0.1132981	0.38 0.1557418		

	PET111::YFP, default cluster					PET111::YFP, cell type				
	Pearson	Spearman	Kendall	P-sctransform	CSCORE	Pearson	Spearman	Kendall	P-sctransform	CSCORE
Count data	0.91 1.15181e-06	0.7 0.00232015	0.57 0.002622326	0.91 7.062342e-07	1 NA	0.88 0.001665684	0.98 4.266322e-06	0.93 0.000541501	0.996 1.710614e-08	1 NA
Normalized	0.89 4.106974e-06	0.66 0.005146845	0.52 0.006231881			0.995 3.133666e-08	0.98 4.266322e-06	0.93 0.000541501		
Scaled	0.89 3.20922e-06	0.66 0.005146845	0.52 0.006231881			0.995 3.661535e-08	0.98 4.266322e-06	0.93 0.000541501		

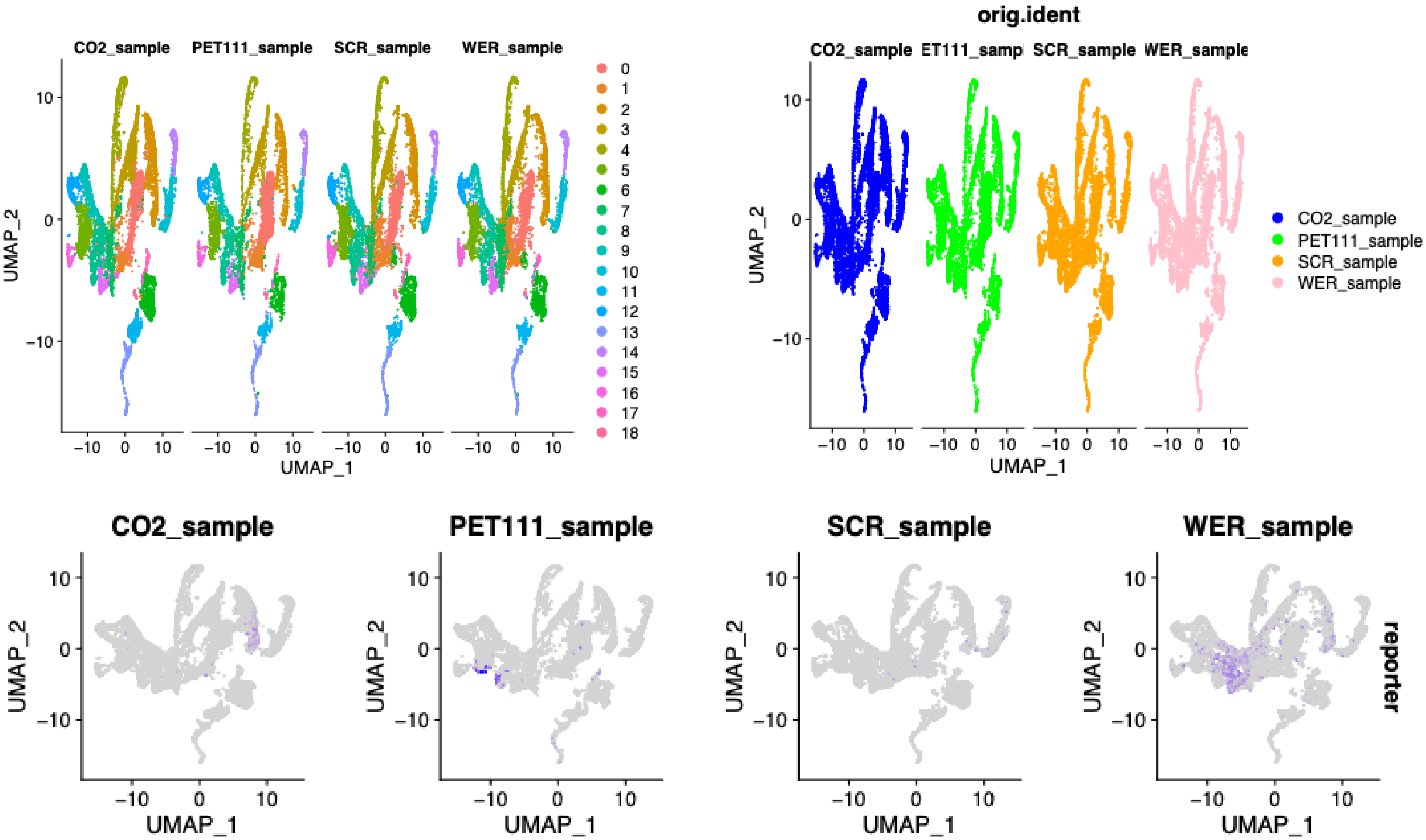
Supplementary Figure 9.

The table presents the correlation values between promoter-reporter genes and their corresponding native genes across different data types (counts, normalized, and scaled) in various samples. Five correlation methods were applied. The top number in each cell represents the correlation strength, while the bottom number indicates the p-value.



Supplementary Figure 10.

The Venn diagrams illustrate the number of cells expressing promoter-reporter genes and their native counterparts in the original data, as well as in the imputed data generated by different methods, across four different samples.



Supplementary Figure 11. Integration across 4 samples