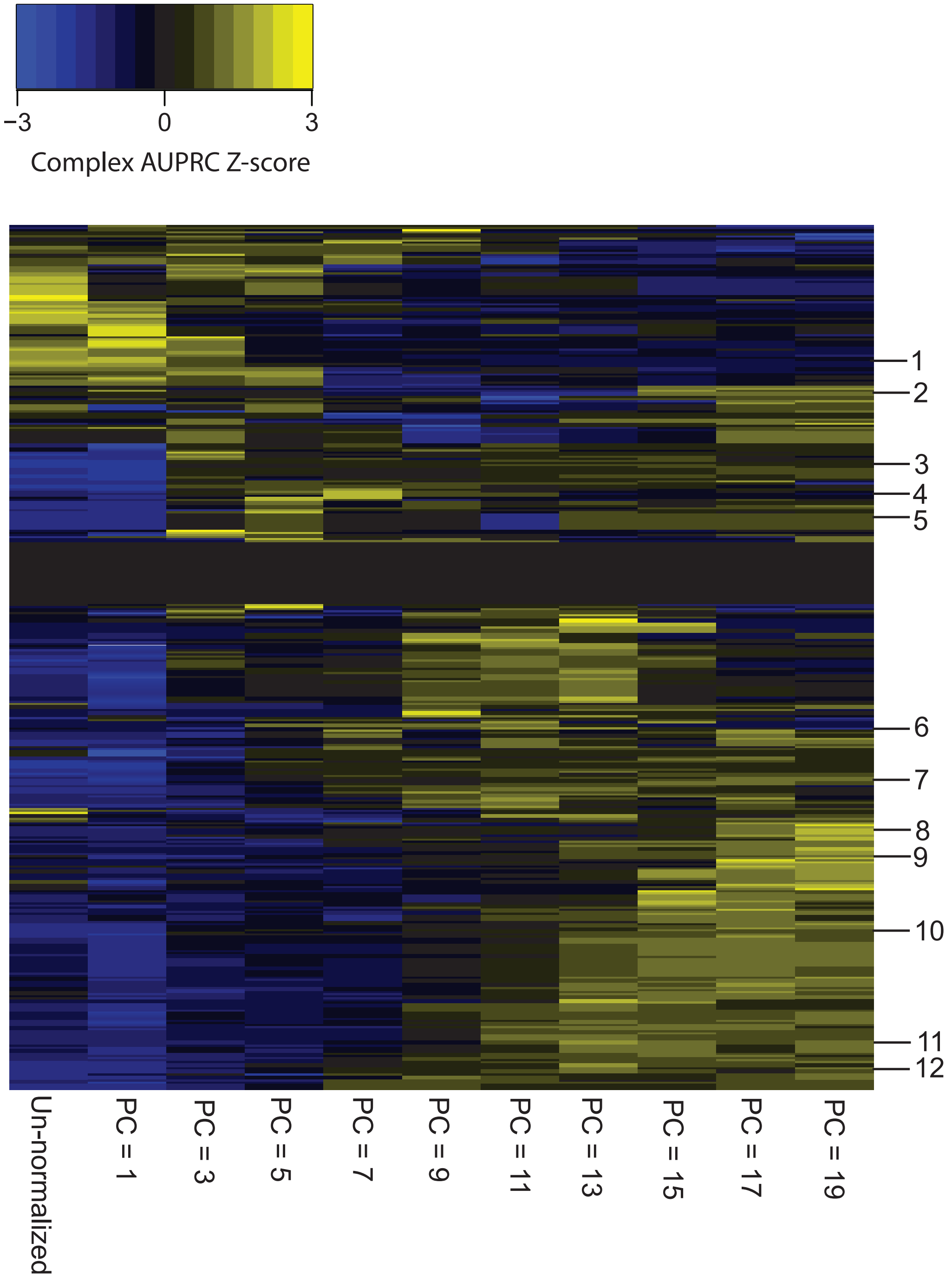
Expanded View Figures

Figure EV1. CORUM complex z-scores of AUPRC values for PCA normalization.

AUPRC values per complex (rows) for the un-normalized DepMap 20Q2 data (Data ref: Broad DepMap, 2020) (far left column) are compared to PCA-normalized DepMap data with the first 1, 3, 5, 7, 9, 11, 13, 15, 17, or 19 PCs removed. AUPRC values have been z-score-normalized per row for visualization purposes.

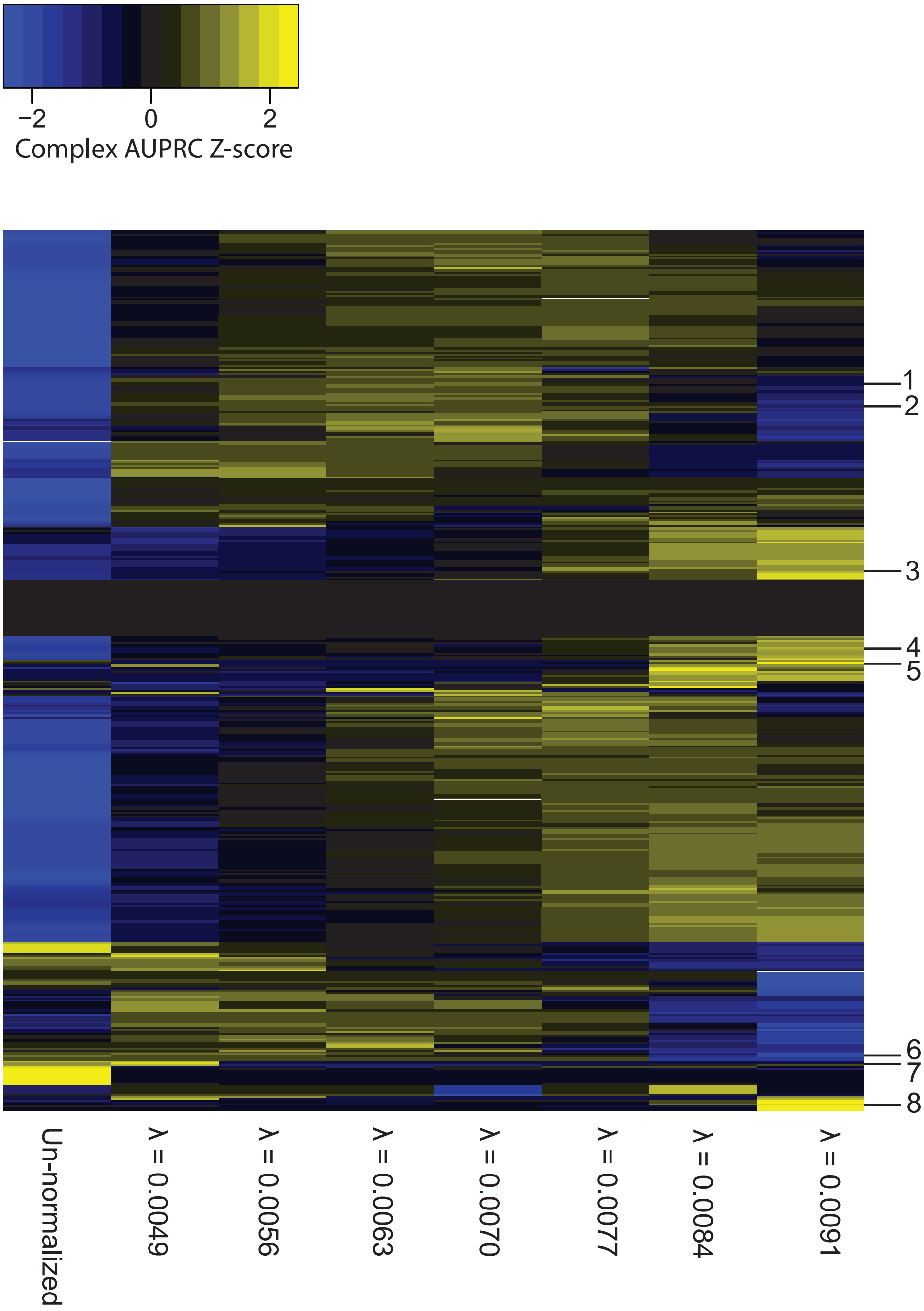


1. 55S ribosome, mitochondrial
2. ORC 2-4 complex
3. WDR20-USP46-UAF1 complex
4. MCM2-MCM6-MCM7 complex
5. RAD50-MRE11-NBN-p200-p350 complex
6. mTOR complex
7. MCM complex
8. CDK8 subcomplex
9. BORC complex
10. KICSTOR complex
11. Condensin I complex
12. MIB complex

Figure EV1.

Figure EV2. CORUM complex z-scores of AUPRC values for robust PCA normalization.

AUPRC values per complex (rows) for the un-normalized DepMap 20Q2 data (Data ref: Broad DepMap, 2020) (far left column) are compared to robust RPCA-normalized DepMap data for $\lambda \approx$ 0.0049, 0.0056, 0.0063, 0.007, 0.0077, 0.0084, 0.0091. AUPRC values have been z-score-normalized per row for visualization purposes.

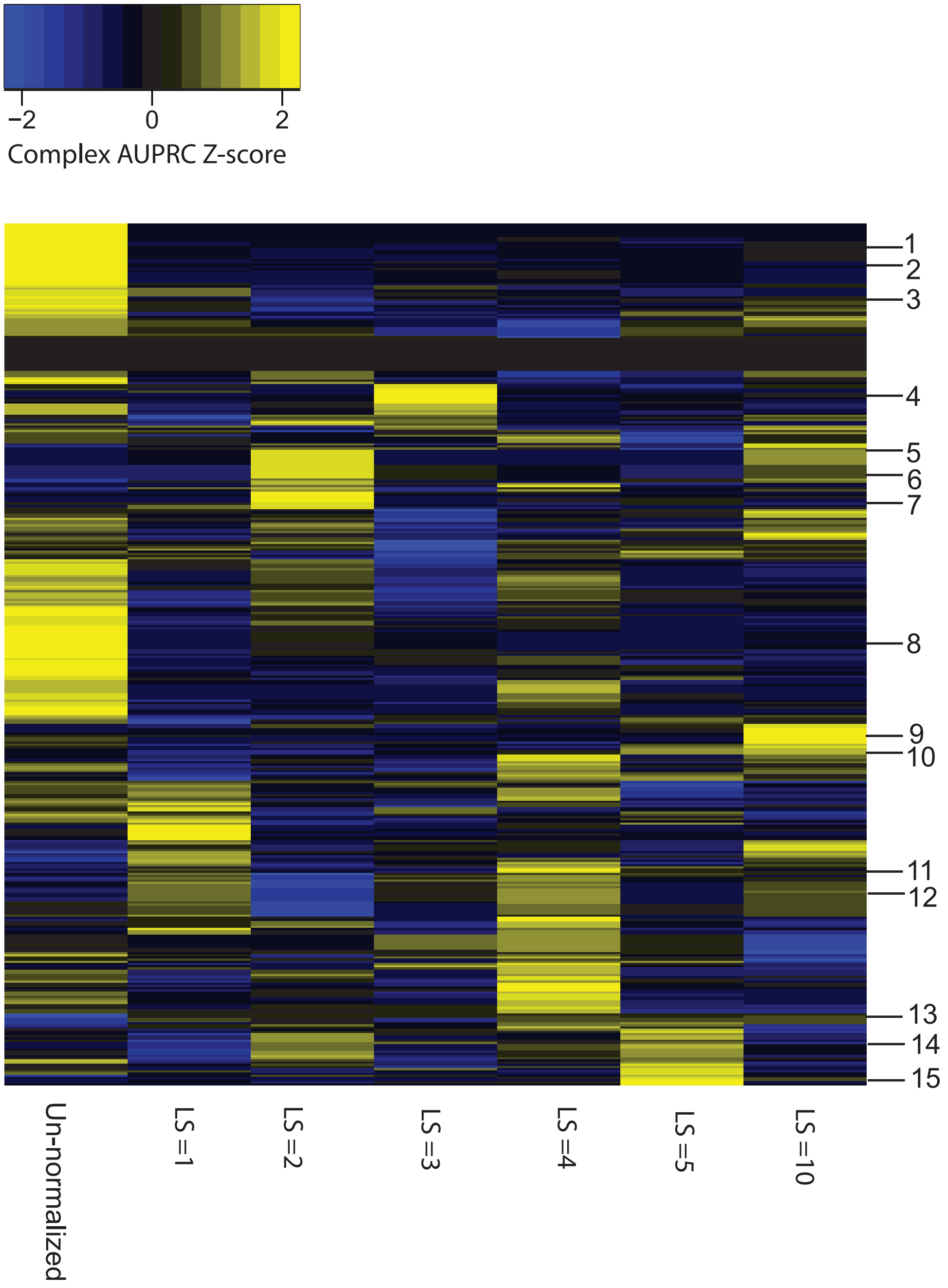


1. ESCRT-II complex
2. ORC 2-4 complex
3. WICH complex
4. RAD17-RFC complex
5. HUSH complex
6. 55S ribosome, mitochondrial
7. hMediator complex
8. KEOPS complex

Figure EV2.

Figure EV3. CORUM complex z-scores of AUPRC values for autoencoder normalization.

AUPRC values per complex (rows) for the un-normalized DepMap 20Q2 data (Data ref: Broad DepMap, 2020) (far left column) compared to AE-normalized DepMap data with latent space size (LS) = 1, 2, 3, 4, 5, 10. AUPRC values have been z-score-normalized per row for visualization purposes.



1. 55S ribosome
2. 39S ribosomal subunit
3. Respiratory chain complex
4. MBD1-MCAF1-SETDB1 complex
5. DNA-PK-Ku complex
6. DNA ligase IV-XRCC4 complex
7. LSm2-8 complex
8. mTOR complex
9. KATNA1-KATNAL1-KATNB1-KATNBL1 complex
10. Chromosomal passenger complex CPC
11. ANCO1-HDAC3 complex
12. BAF complex
13. Wave-2 complex
14. PAC3-PAC4 complex
15. FA complex

Figure EV3.

Figure EV4. *K*-nearest neighbor classifier tissue-lineage prediction results comparing un-normalized and RPCO-normalized DepMap 20Q2 (Data ref: Broad DepMap, 2020) cell line similarity networks.

A Overall F1, Precision, and Recall scores (weighted mean across classes) across different values of K (x-axis). Dashed line represents the scores from a baseline random classifier.

B Class-level (tissue-lineage) F1, Precision, and Recall scores for $K = 5$ where x-axis depicts tissue lineage.

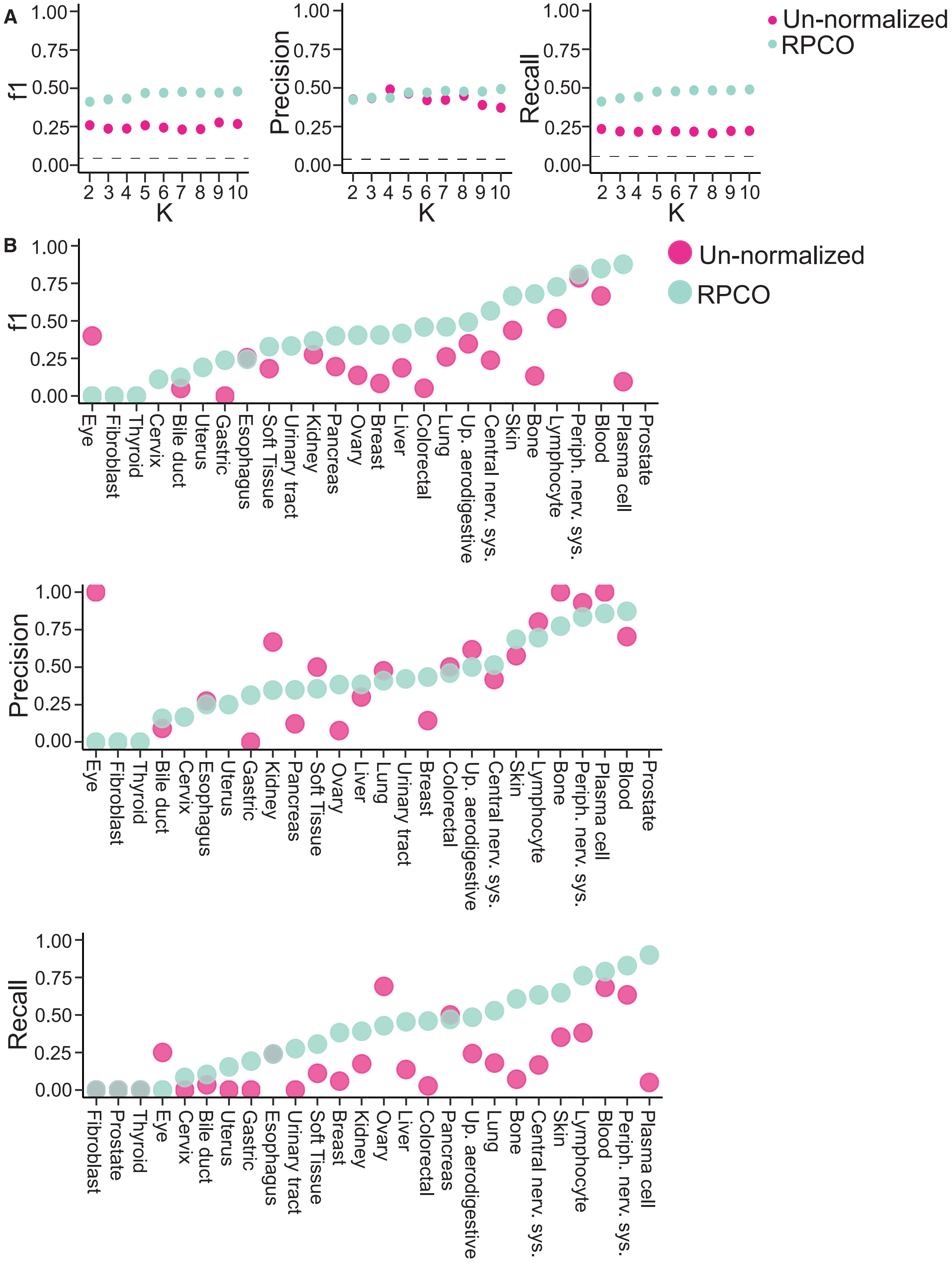

A
f1
Precision
Recall
1.00
0.75
0.50
0.25
0.00
2 3 4 5 6 7 8 9 10
K
Un-normalized
RPCO
B
f1
Eye
Fibroblast
Thyroid
Cervix
Bile duct
Uterus
Gastric
Esophagus
Soft Tissue
Urinary tract
Kidney
Pancreas
Ovary
Breast
Liver
Colorectal
Lung
Up. aerodigestive
Central nerv. sys.
Skin
Bone
Lymphocyte
Periph. nerv. sys.
Blood
Plasma cell
Prostate
Precision
Recall

Figure EV4.

Figure EV5. Confusion matrix from *K*-nearest neighbor classifier (K = 5) showing prediction results for each class (tissue lineage) of cell lines from un-normalized and RPCO-normalized DepMap 20Q2 (Data ref: Broad DepMap, 2020) cell line similarity networks.

X- and y-axes are class labels. The top row depicts the true number of cell lines in each class, and the right-most column is the number of predicted cell lines for each class. The diagonal represents true positives. Each row represents false positives, and each column represents false negatives sans the diagonal entry.

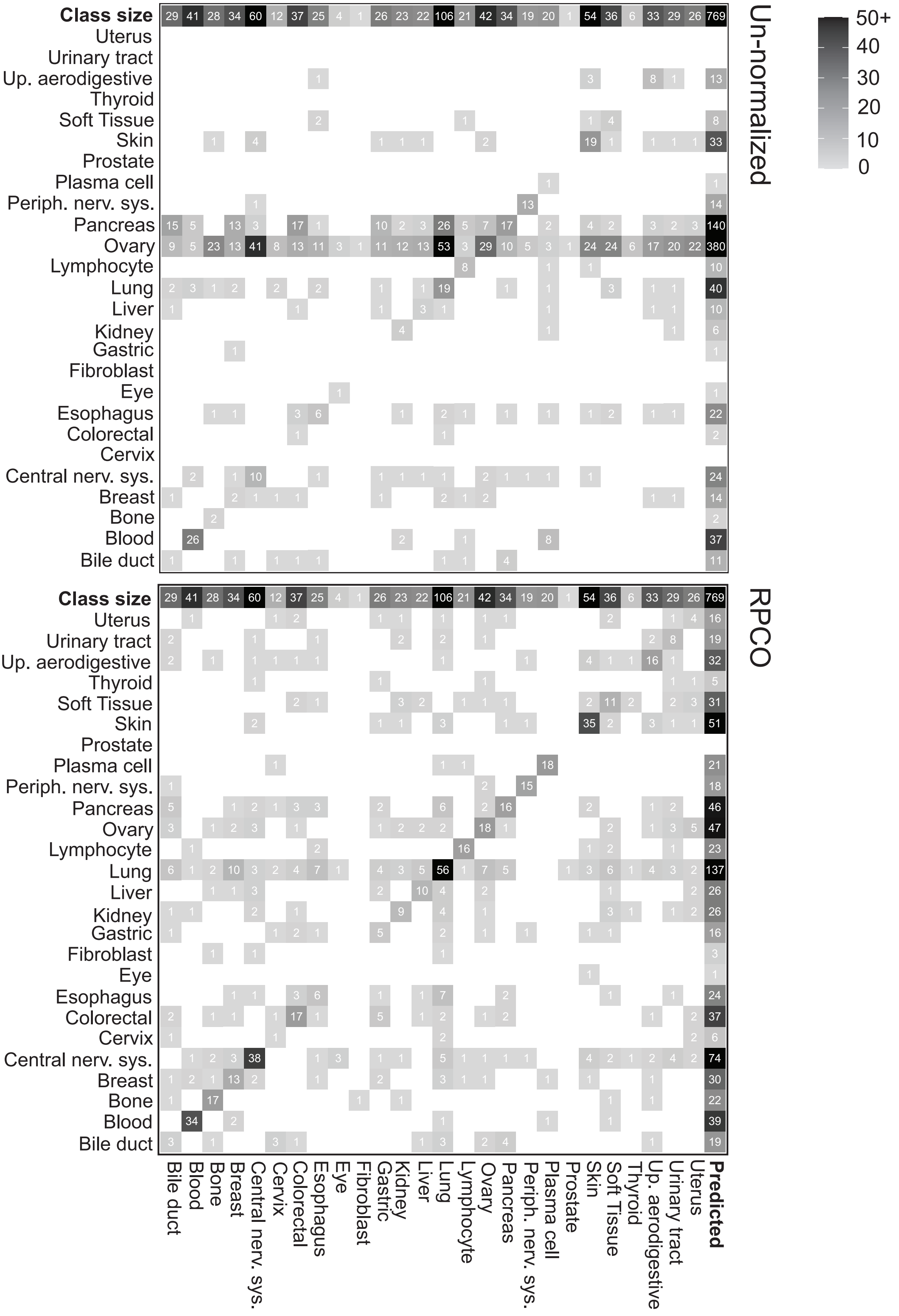

Un-normalized
RPCO
Class size
Uterus
Urinary tract
Up. aerodigestive
Thyroid
Soft Tissue
Skin
Prostate
Plasma cell
Periph. nerv. sys.
Pancreas
Ovary
Lymphocyte
Lung
Liver
Kidney
Gastric
Fibroblast
Eye
Esophagus
Colorectal
Cervix
Central nerv. sys.
Breast
Bone
Blood
Bile duct
Predicted
29 41 28 34 60 12 37 25 4 1 26 23 22 106 21 42 34 19 20 1 54 36 6 33 29 26 769
50+
40
30
20
10
0

Figure EV5.