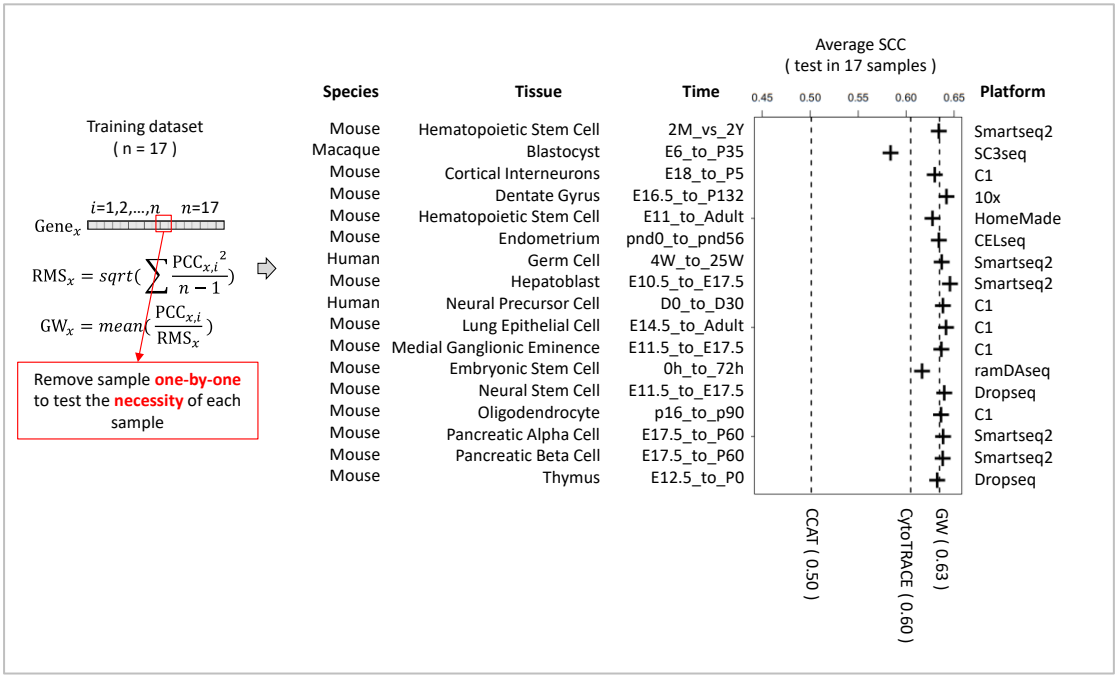
Supplementary Figures

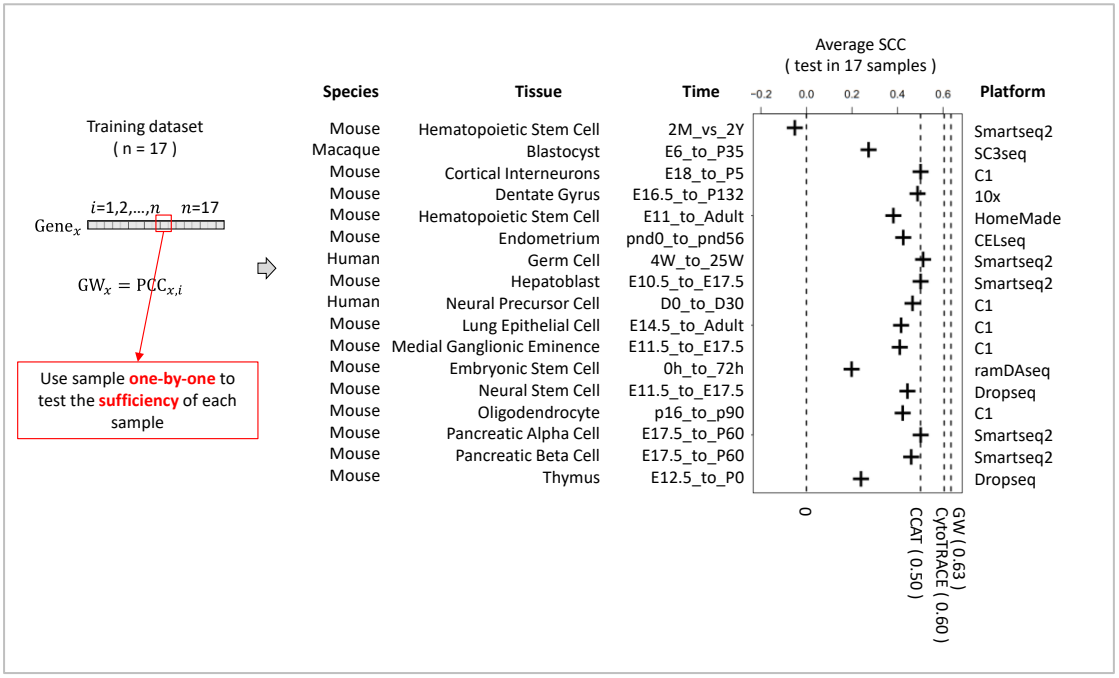
Supplementary Figure 1. Screening the necessity of each sample

The workflow of checking the necessity of each sample in the training dataset.



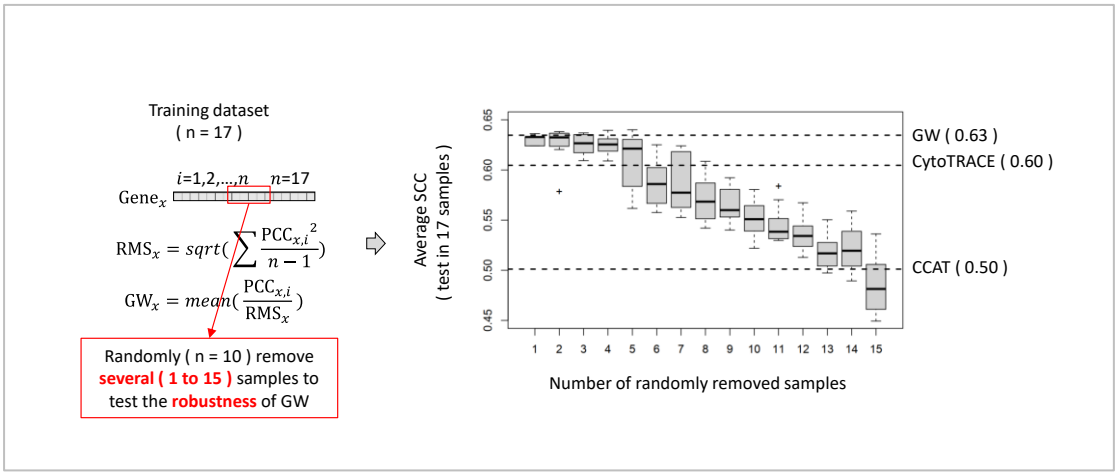
Supplementary Figure 2. Screening the sufficiency of each sample

The workflow of checking the sufficiency of each sample in the training dataset.



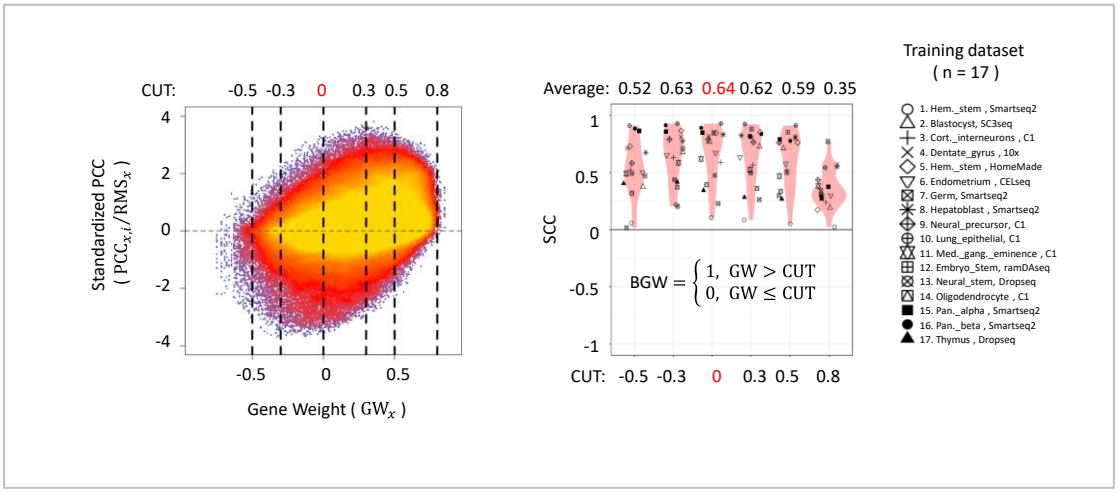
Supplementary Figure 3. Testing the robustness of GW

The workflow of checking the robustness of our GW.



Supplementary Figure 4. Screening the thresholds of BGW

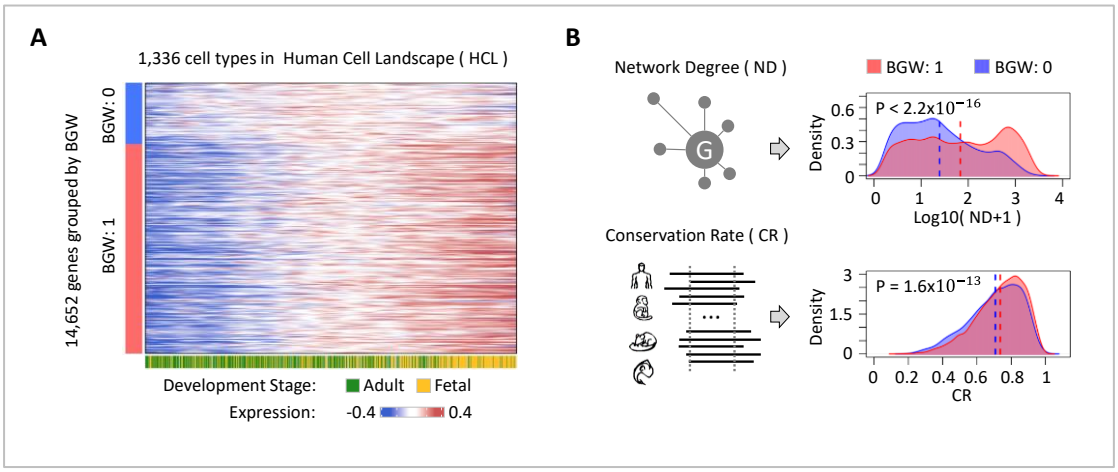
The performance of BGW at different thresholds.



Supplementary Figure 5. The application of BGW with human cell types in HCL and the biological characteristics of BGW

(A) The cell types of HCL are sorted by the DP inferred by using BGW.

(B) The biological characteristics of BGW. In addition to testing the performance of BGW, we then try to reveal the biological characteristics of BGW. Firstly, we use the network, named "net17Jan16", provided by CCAT to calculate the network degree of each gene. Genes with positive GW tend to have a significantly higher network degree than genes with negative GW (P-value<2.2×10^{-16}), demonstrating that genes with positive GW may contribute more to the cell's entropy [1, 2] than genes with negative GW. Secondly, we calculate a conservation rate for each gene by counting consistent amino acids of 100 species (UCSC, hg19, 100-way). We find that genes with positive GW tend to have a significantly higher conservation rate than genes with negative GW (P-value=1.6×10^{-13}), which can be associated with the findings reported by Cardoso-Moreira *et al.* [3]. In 2019, Cardoso-Moreira *et al.* reported that transcription profiles of different organs and species are more similar at early stages (*i.e.*, early organogenesis), while the expression levels of newly generated species-specific genes increase at late stages (*i.e.*, adulthood).



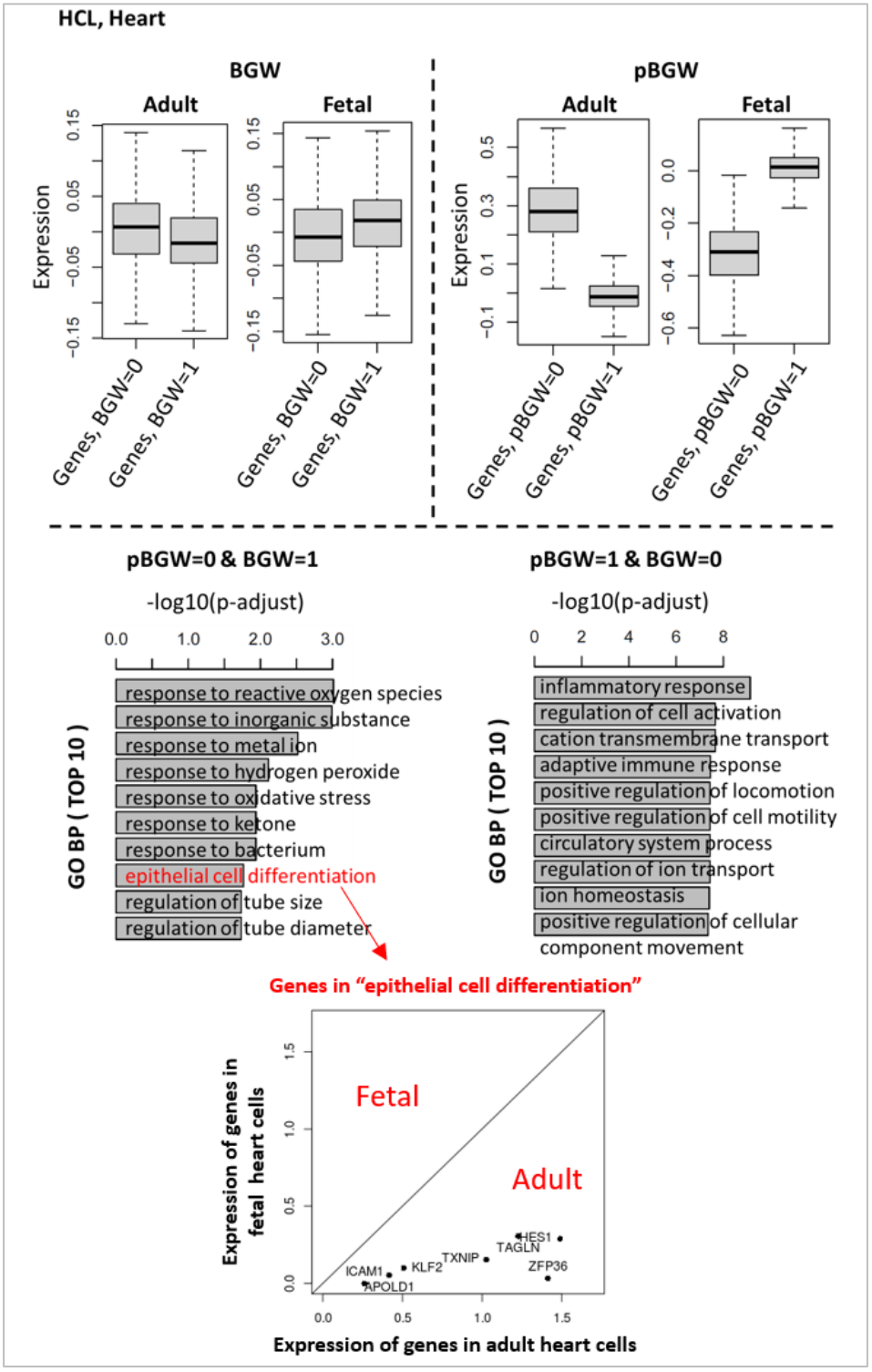
Supplementary Figure 6. BGW is applicable in tissues that are not covered by the training dataset.

We use the novel tissues (not covered by training dataset) in HCL to show that our BGW is able to identify fetal cell types from adult cell types, suggesting that BGW contributes to the developmental potential of the biological system/tissues not mentioned in the training dataset.

Supplementary Figure 7. The difference between BGW and predicted BGW (pBGWs).

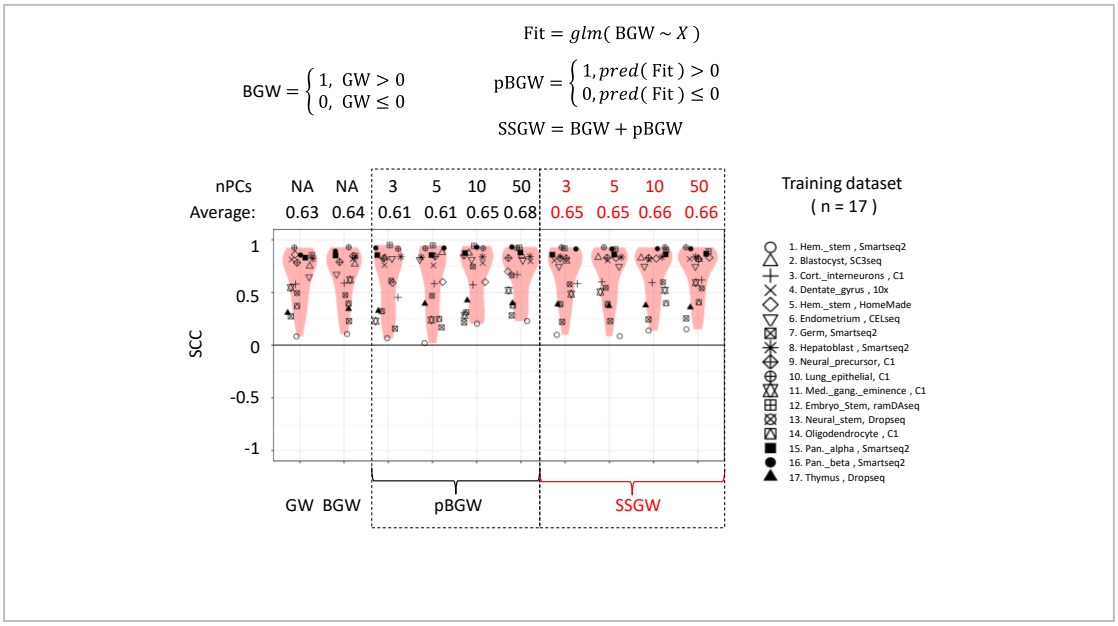
We use a heart sample in HCL to demonstrate and compare the meanings of BGW and pBGW. The expression of genes with BGW=1 (pBGW=1) is higher than genes with BGW=0 (pBGW=0) in the cells of fetal samples, while the expression of genes with BGW=0 (pBGW=0) is higher than genes with BGW=1 (pBGW=1) in the cells of adult samples, demonstrating that genes with positive BGW (pBGW) are a group of genes that are activated in the early developmental stage (fetal), while genes with zero BGW (pBGW) are a group of genes that are activated in the late developmental stage (adult). Furthermore, we have noticed that pBGW shows a clearer pattern than BGW, suggesting that the sample-specific information can help pBGW to identify more accurate gene groups. To further compare the functional differences of pBGW and BGW, we use genes with “pBGW=0 & BGW=1” and genes with “pBGW=1 & BGW=0” to conduct

gene set enrichment analysis (GO BP). We find that genes with "pBGW=0 & BGW=1" are enriched in a term named "epithelial cell differentiation". There are seven genes that are involved in this pathway: *APOLD1*, *HES1*, *ICAM1*, *KLF2*, *TAGLN*, *TXNIP*, and *ZFP36*. In this figure, we have shown that all of those seven genes are highly expressed in adult heart cells but lowly expressed in fetal heart cells, suggesting that pBGW can be more suitable for accurately inferring the DP in this sample.



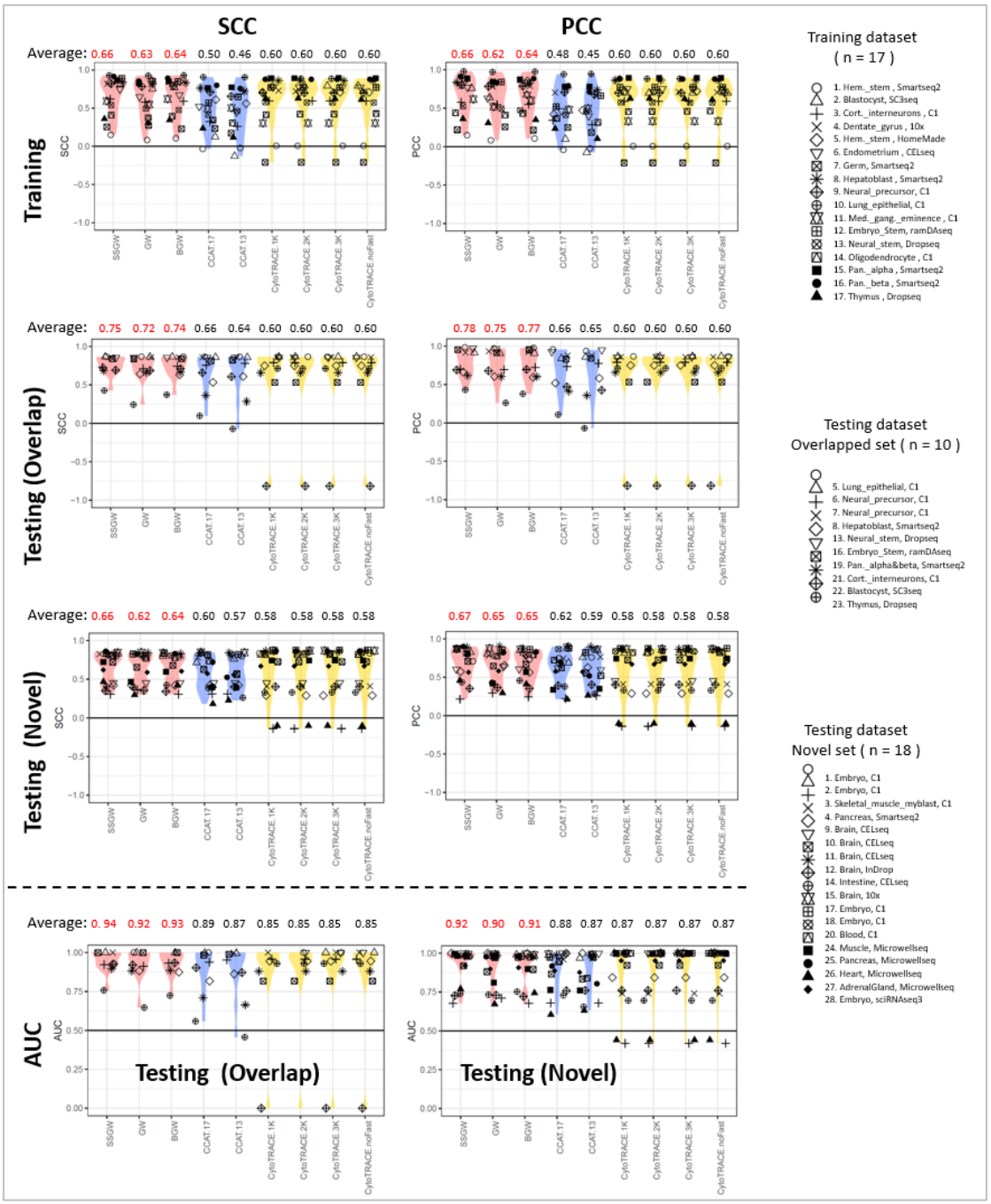
Supplementary Figure 8. The comparison of GW, BGW, pBGW, and SSGW using different numbers of PCs

"nPCs" stands for the number of used PCs. This supplementary figure is related to **Figure 2D**



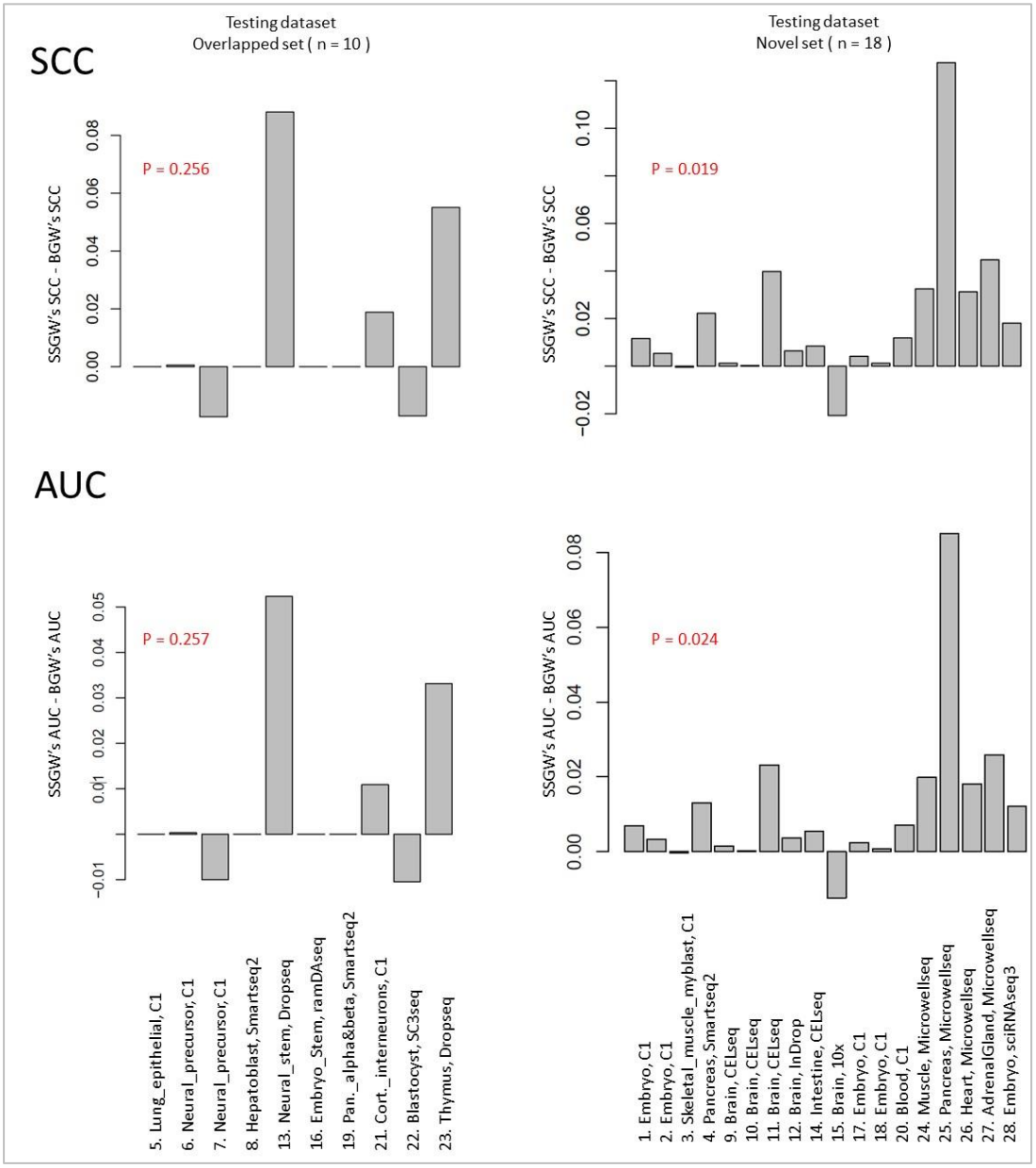
Supplementary Figure 9. Parameter tuning of CCAT and CytoTRACE

CCAT has one parameter: the gene network. We use two different networks ("net17Jan16.m" and "net13Jun12.m") provided by the author of CCAT to test the performance of CCAT. CytoTRACE has one parameter: the down-sampling size. We use "1k", "2k", "3k", and "noFast (no down-sampling)" to test the performance of CytoTRACE. "SSGW" stands for FitDevo. This supplementary figure is related to **Figure 2D, Figure 3C,** and **Figure 3D.**



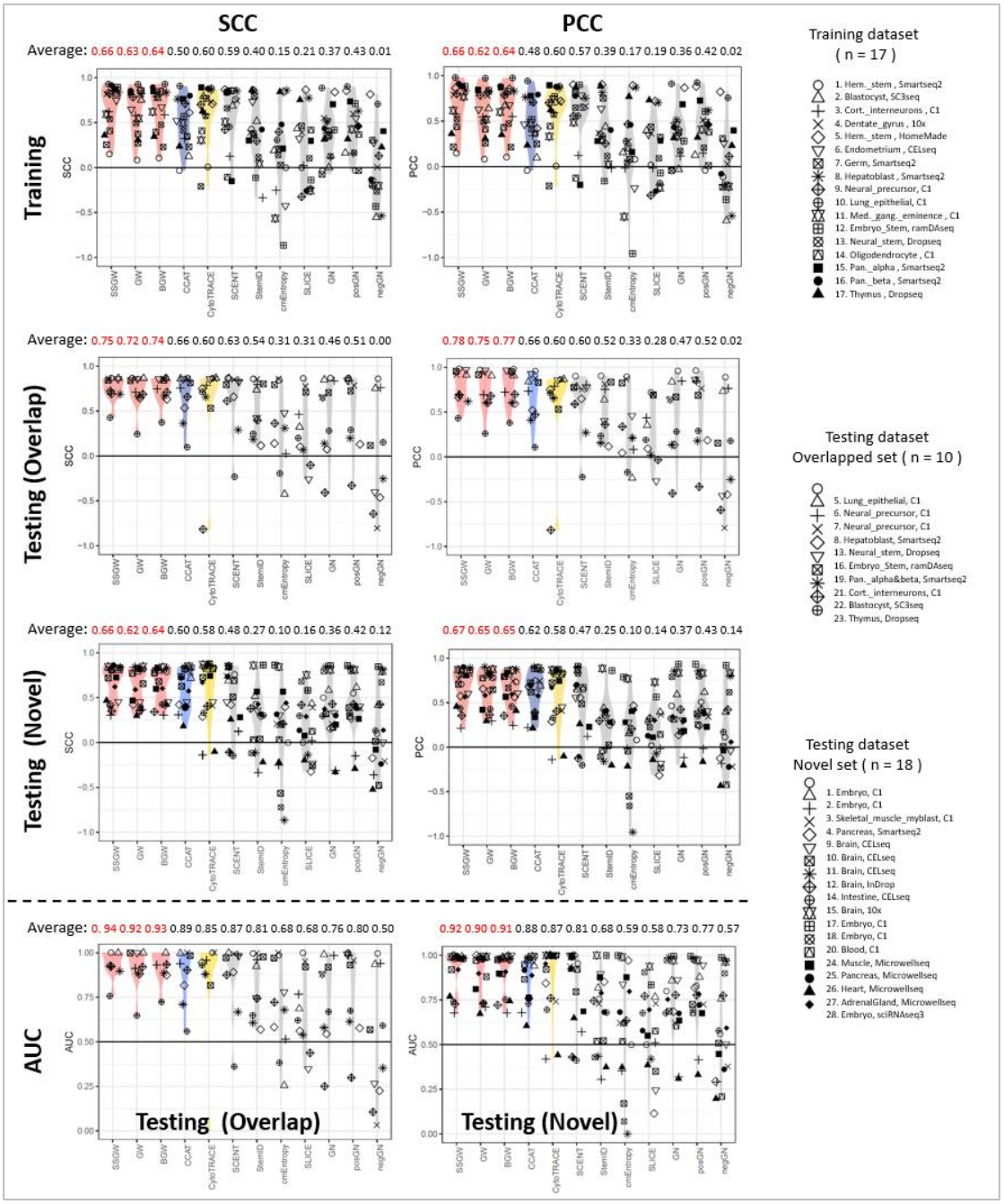
Supplementary Figure 10. The detailed difference between the performance of SSGW and BGW with the testing dataset

This supplementary figure is related to **Figure 3C** and **Figure 3D.**



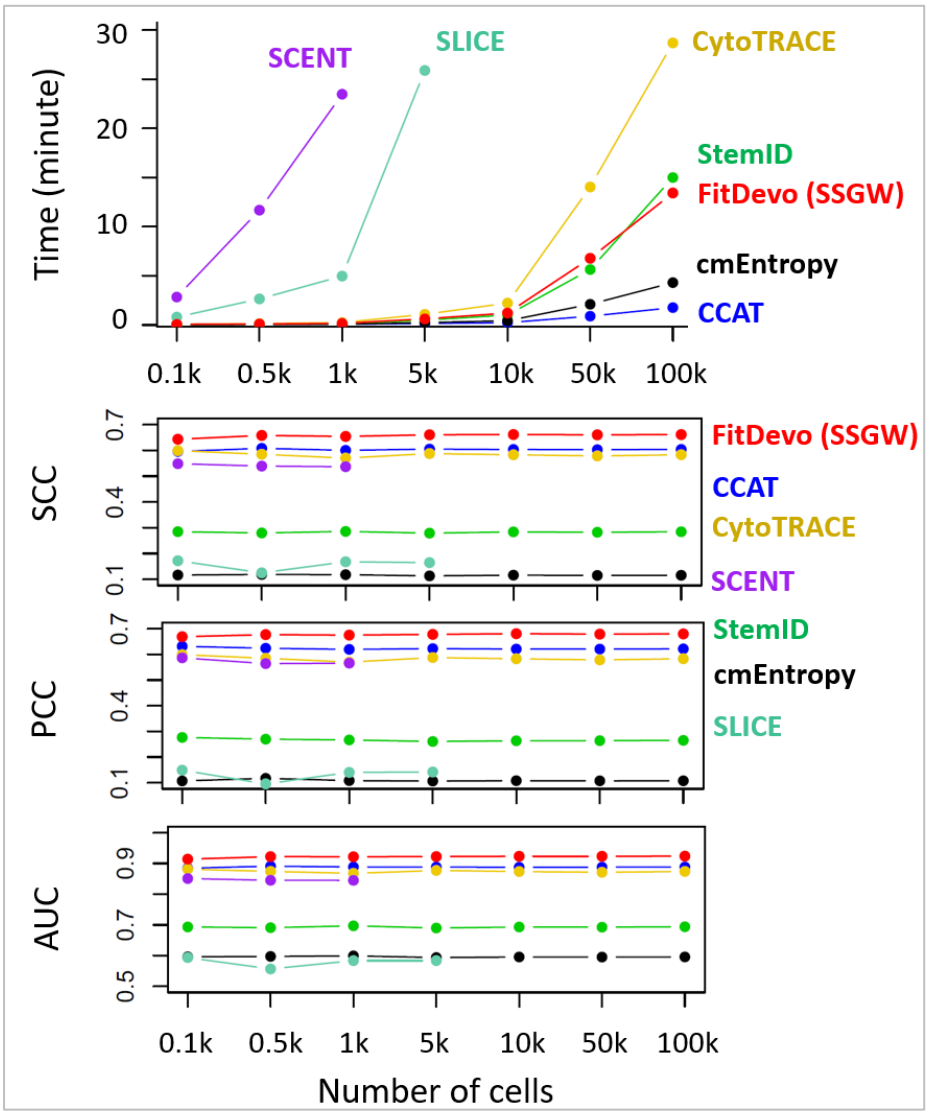
Supplementary Figure 11. The performance measured by SCC, PCC, and AUC

Following the study of CCAT, we have downloaded and run those competing methods (CCAT, CytoTRACE, SCENT, StemID, cmEntropy, and SLICE). scEnergy is not tested here, because we don't have its operating platform (MATLAB). However, our competing results of other methods are similar with the results described in the paper of CCAT [2], and the study of CCAT shows that the performance of scEnergy is close to StemID. "SSGW" stands for FitDevo. This supplementary figure is related to **Figure 2D, Figure 3C,** and **Figure 3D.**



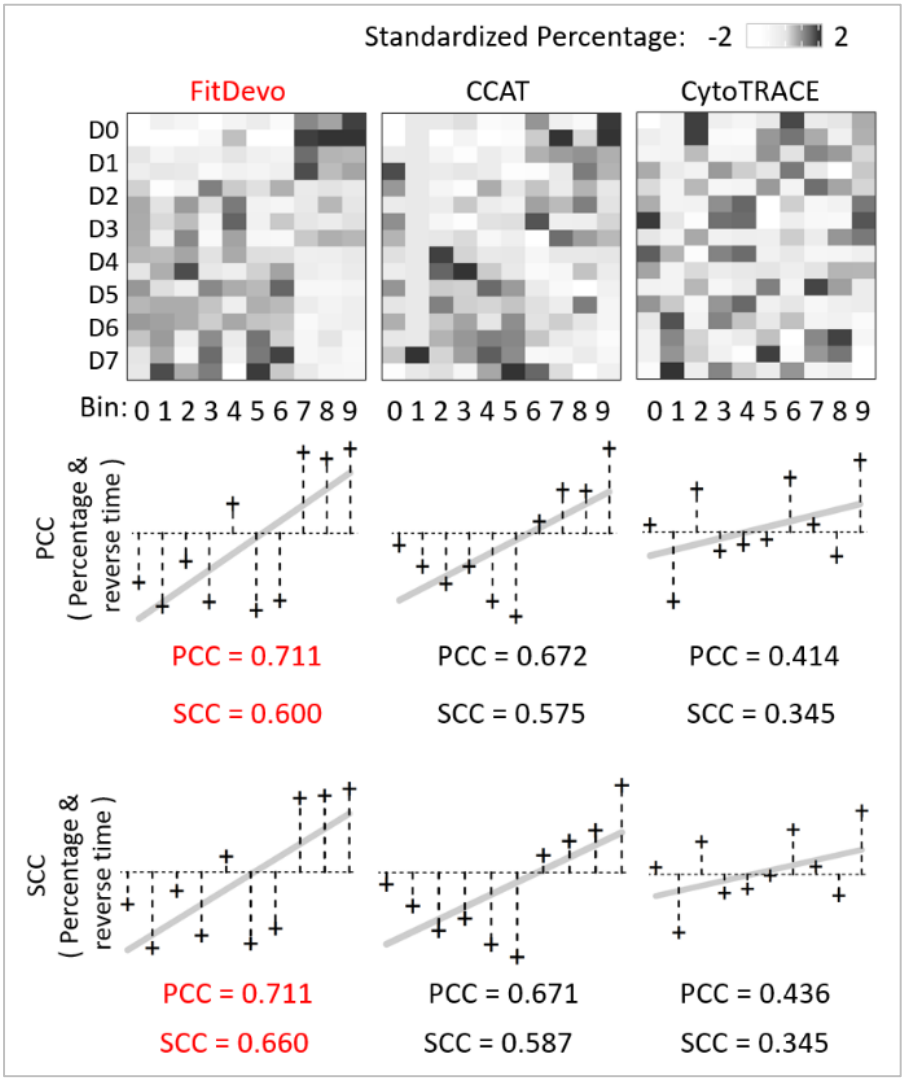
Supplementary Figure 12. The computational scalability evaluation

We use the 18 novel samples in our testing dataset to conduct this computational scalability evaluation. We randomly extract cells from each sample to build a pseudo-sample with a large number of cells (*e.g.*, 50k, 100k, *etc.*). The running time (minute) is the average running time of the 18 samples. When dealing with 10k and 100k cells, FitDevo will cost around 18G and 134G memory (1 core), respectively.



Supplementary Figure 13. Using PCC and SCC to evaluate the results of deconvolution analysis

This supplementary figure is related to **Figure 4D.**

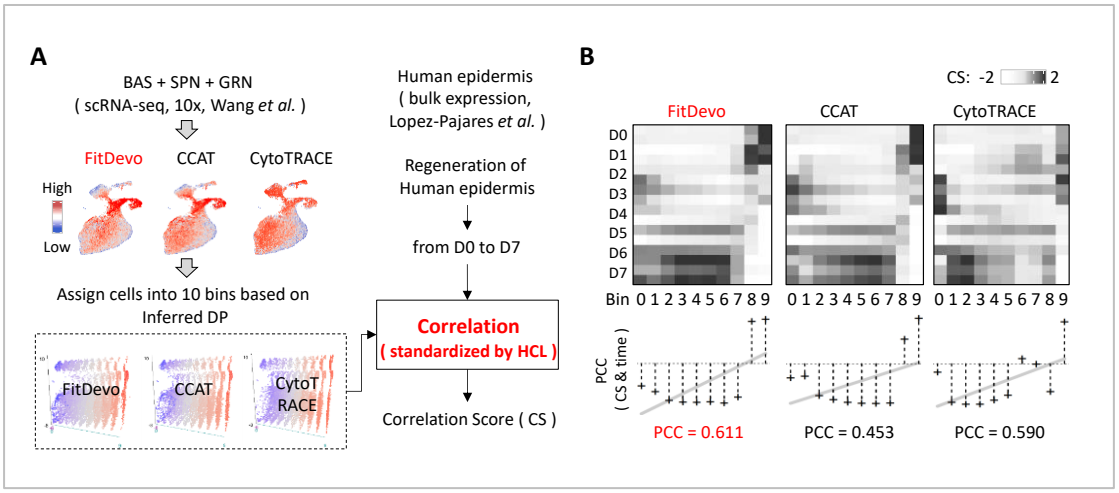


Supplementary Figure 14. The deconvolution analysis of epidermis

(A) The workflow of conducting deconvolution analysis using a correlation-based method. In this study, we have used two methods to conduct deconvolution analysis: CIBERSORTx [4] and a correlation-based method. For CIBERSORTx, we first use the default parameters of CIBERSORTx to create signature matrix. Because the computational efficiency is low when dealing with tens of thousands of single cells, we have used down-sampling to randomly choose 1,000 cells from 14,177 epidermal cells. Notably, we have tried different random seeds and have got similar results, which are not shown. Then, we upload the bulk data to the server, and use "S-mode" to conduct batch correction. Finally, we get a cell-type fraction matrix. Considering CIBERSORTx has many adjustable parameters, we therefore use a more straightforward way (correlation-based method) to conduct deconvolution analysis. Firstly, for each bin in scRNA-seq data, we average gene expression value to get a single-cell reference (without down-sampling). Secondly, we use ComBat [5] to remove the batch effect among the single-cell reference, bulk data, and HCL data. Thirdly, we calculate PCC between the bulk data and the single-cell reference. Fourthly, we calculate PCC between the bulk data and the HCL data. Finally, for each sample of the bulk data, we put all PCCs together (the PCCs of the single-cell reference and the HCL data) and conduct standardization (mean is 0, standard deviation is 1) to get the final correlation scores of the single-cell bins.

(B) The results of correlation-based deconvolution analysis.

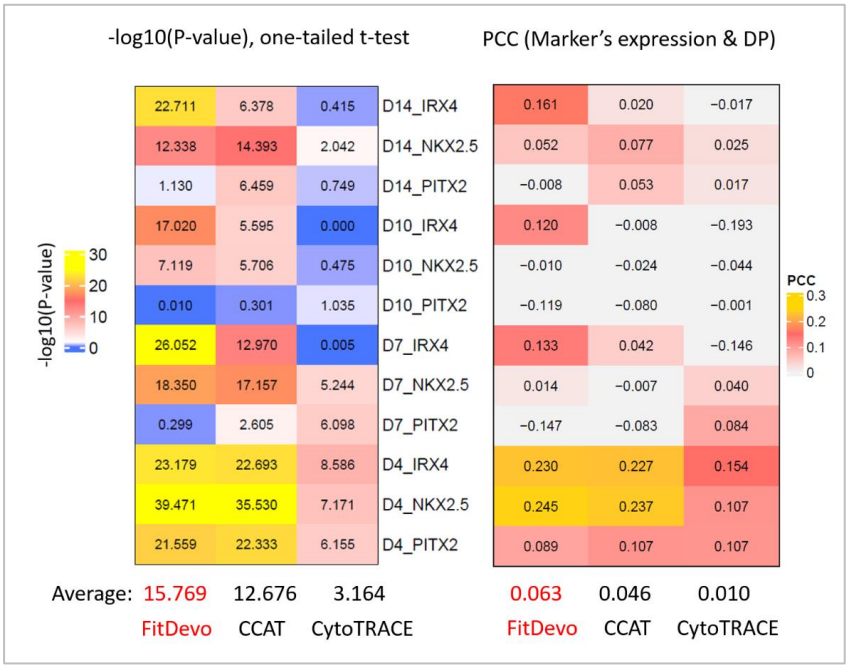
This supplementary figure is related to **Figure 4C** and **Figure 4D.**



Supplementary Figure 15. The spatial transcriptomic data analysis of immature myocardial markers during chick heart development.

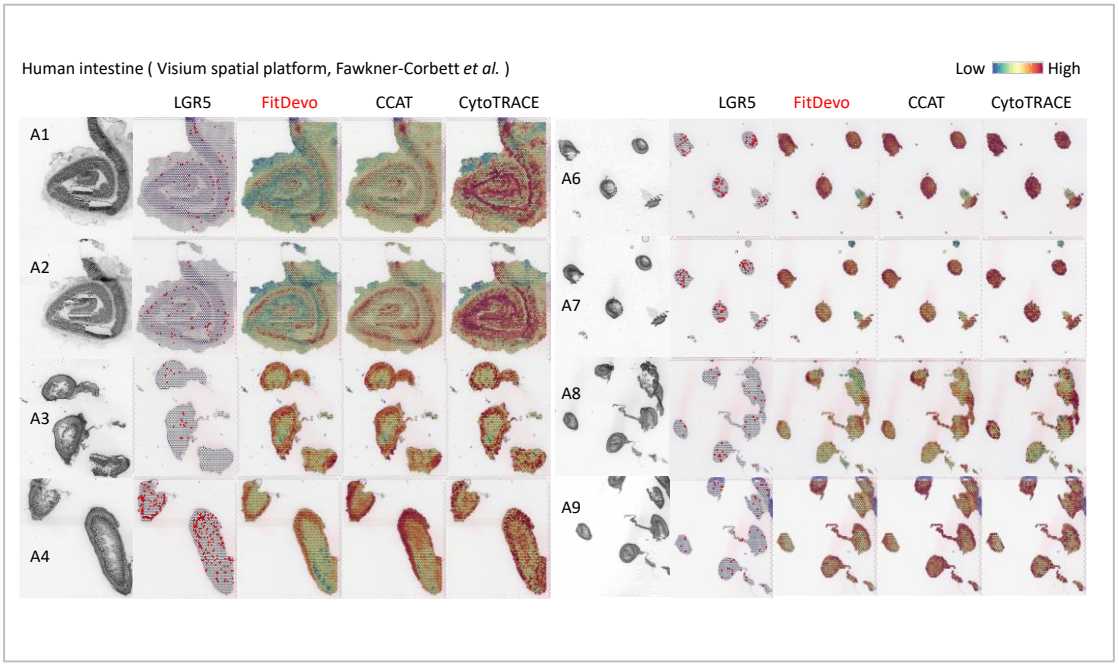
In this figure, we provide the results of *NKX2-5*, *PITX2*, and *IRX4* in D4, D7, D10, and D14 samples. For the comparison of DP between marker positive cells and negative cells, FitDevo has a larger averaged -log10(P-value) (15.769) than CCAT (12.676) and CytoTRACE (3.164). For the PCC between marker's expression and DP, FitDevo consistently has a larger averaged PCC (0.063) than CCAT (0.046) and CytoTRACE (0.010).

This supplementary figure is related to **Figure 5B** and **Figure 5C.**



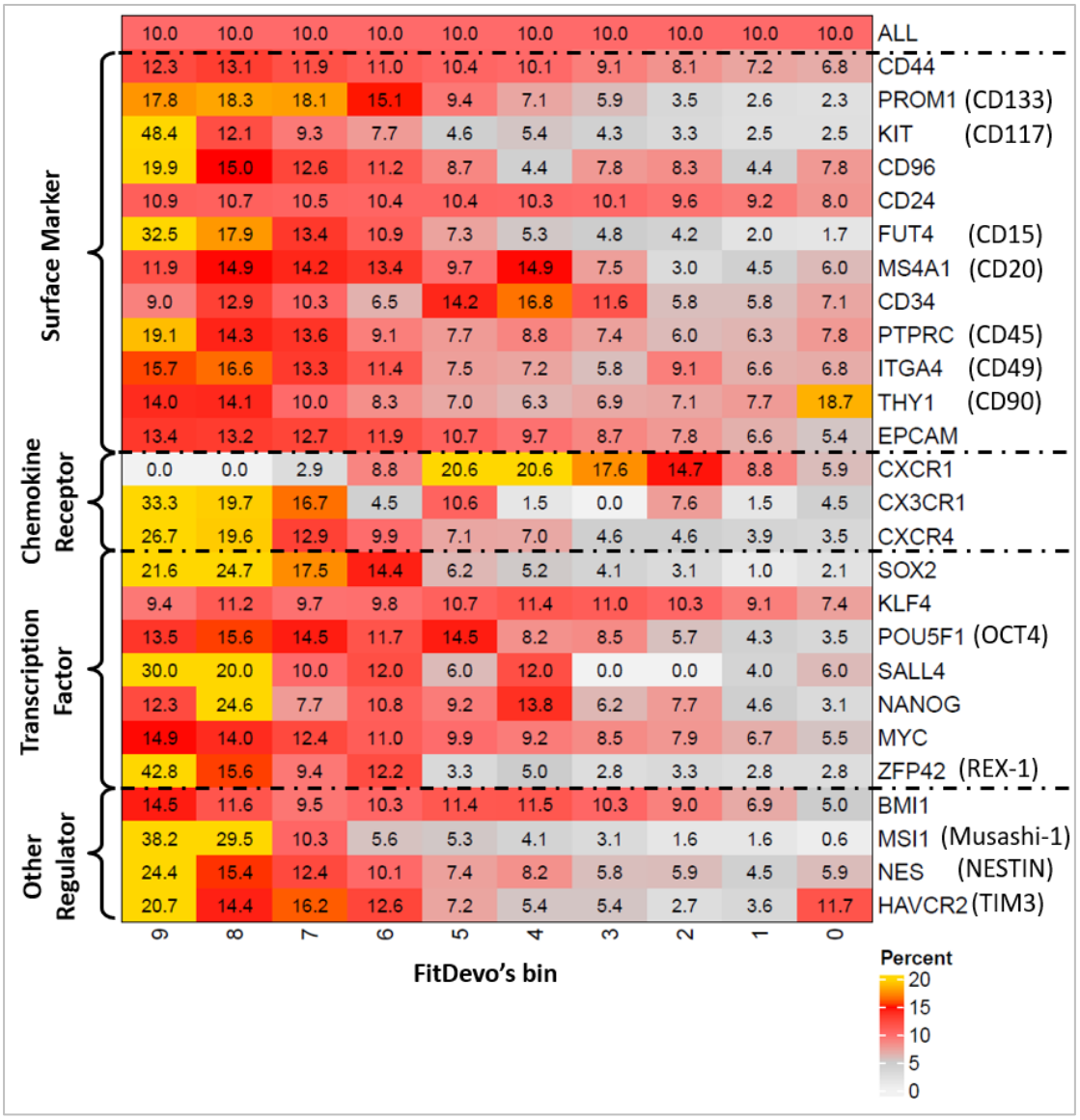
Supplementary Figure 16. The results of all human intestine spatial samples

This supplementary figure is related to **Figure 6A.**



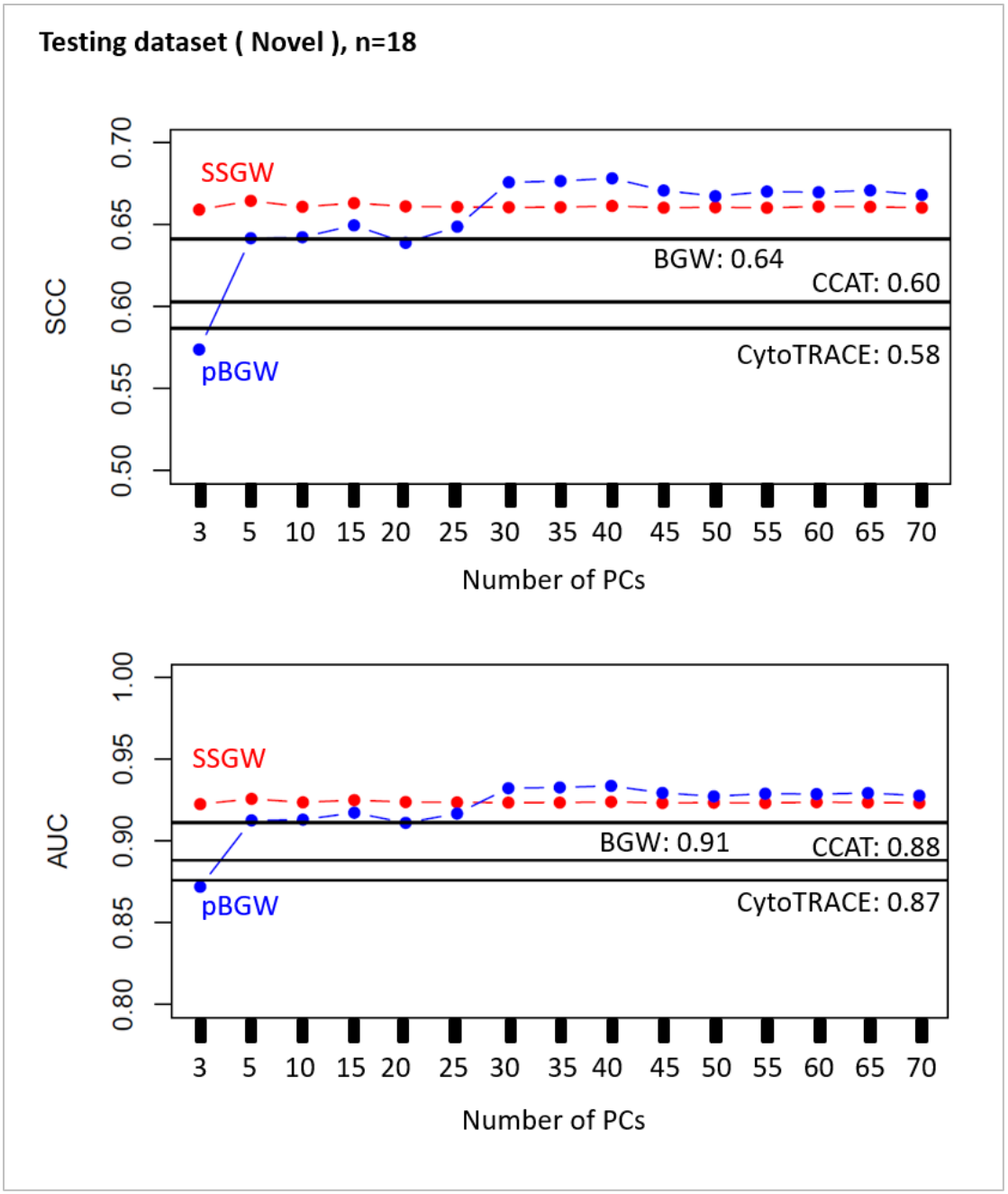
Supplementary Figure 17. Results of different markers

The percentage of the signature (*e.g.*, *CD44*, *PROM1*, and *KIT*, *etc.*) positive cells in each bin of FitDevo. "ALL" indicates the percentage of all cells in each bin. We have collected all 26 signatures (*e.g.*, *SOX2*, *NANOG*, and *POU5F1*, *etc.*) presented in the paper named "*Stemness-Related Markers in Cancer*" [6]. There are four groups of signatures: surface marker, chemokine receptor, transcription factor, and other regulator. For most signatures (22 out of 26), bins with large serial numbers (bin7, bin8, and bin9) have significantly more cells expressing those signatures than those bins with small serial numbers (bin0, bin1, and bin2). This supplementary figure is related to **Figure 7B.**



Supplementary Figure 18. Results of FitDevo when using different number of PCs

We use the 18 novel samples in our testing dataset to conduct this test. The performance of FitDevo (SSGW) is almost independent of the number of used PCs. However, the performance of pBGW is still related to the number of used PCs, suggesting that the instability of pBGW can be caused by the insufficient selection of PCs. The performance will be stable when using more than 30 PCs. Therefore, we suggest users use at least 30 PCs (the default number of calculated PCs in Seurat is 50, which is enough for using FitDevo).



Supplementary Figure 19. A summarization table to summarize the advantage and disadvantage of FitDevo and other known methods

We use the 18 novel samples in our testing dataset to summarize the AUC (**Supplementary Figure 11**). Details about running time are provided at **Supplementary Figure 12**.

Tools	Performance		Practical scenarios			Interface	
	Discrimination Accuracy (DP)	Running Time (100k cells)	Deconvolution analysis	Spatial data analysis	Cancer DP analysis	Environment	Availability
FitDevo	AUC>0.9	10m~30m	Applicable	Applicable	Applicable	R	Github
CCAT	0.9>AUC>0.8	<10m	Moderate	Moderate	Moderate	R	Github
CytoTRACE	0.9>AUC>0.8	10m~30m	Weak	Weak	Weak	R	Github
SCENT	0.9>AUC>0.8	>30m	Not tested (we only test the top three methods with the highest AUC)			R	Github
StemID	AUC<0.8	10m~30m				R	Github
cmEntropy	AUC<0.8	<10m				R	Github
SLICE	AUC<0.8	>30m				R	Author's website

Legend: Excellent | Moderate | Weak

References

1. Teschendorff, A.E. and T. Enver, *Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome.* Nat Commun, 2017. **8**: p. 15599.
2. Teschendorff, A.E., et al., *Ultra-fast scalable estimation of single-cell differentiation potency from scRNA-Seq data.* Bioinformatics, 2021. **37**(11): p. 1528-1534.
3. Cardoso-Moreira, M., et al., *Gene expression across mammalian organ development.* Nature, 2019. **571**(7766): p. 505-509.
4. Newman, A.M., et al., *Determining cell type abundance and expression from bulk tissues with digital cytometry.* Nat Biotechnol, 2019. **37**(7): p. 773-782.
5. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-27.
6. Zhao, W., Y. Li, and X. Zhang, *Stemness-Related Markers in Cancer.* Cancer Transl Med, 2017. **3**(3): p. 87-95.