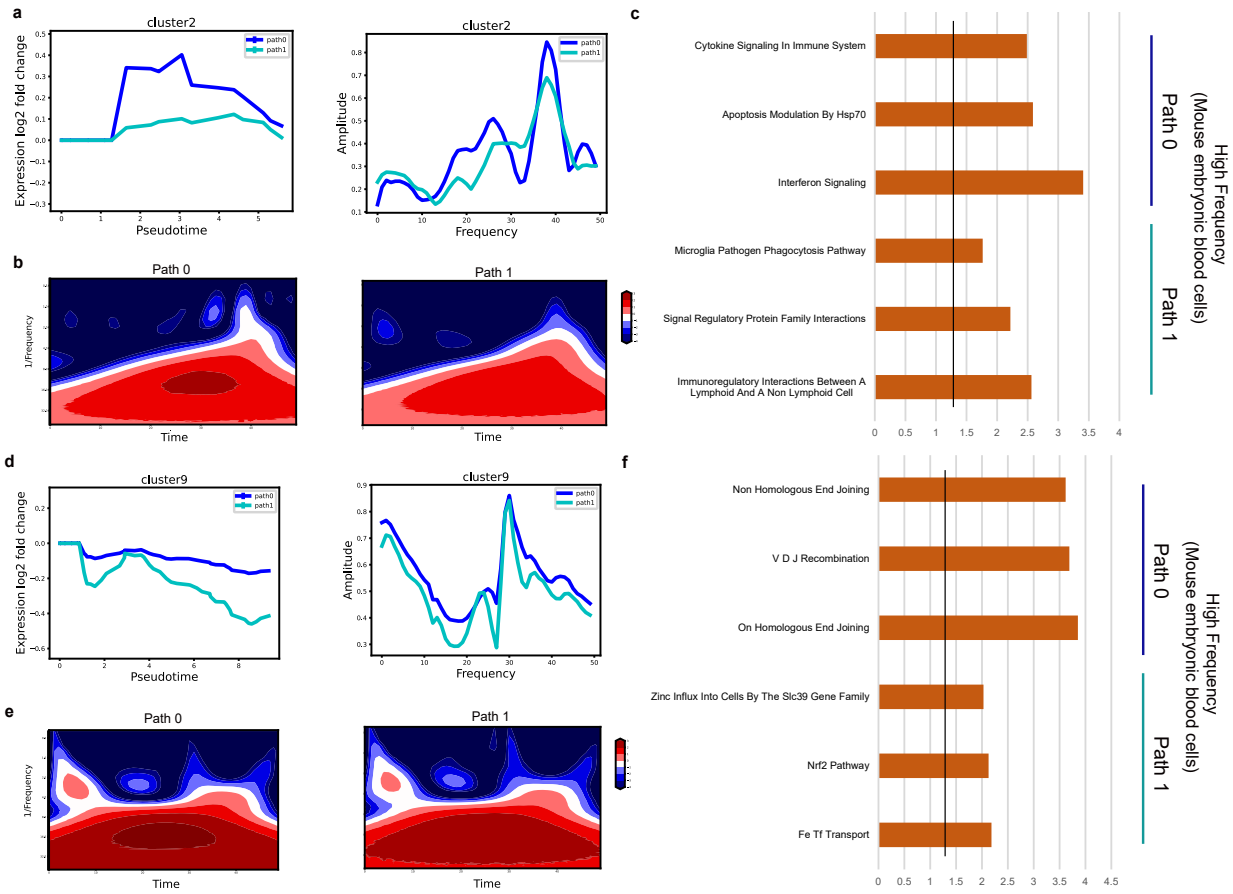
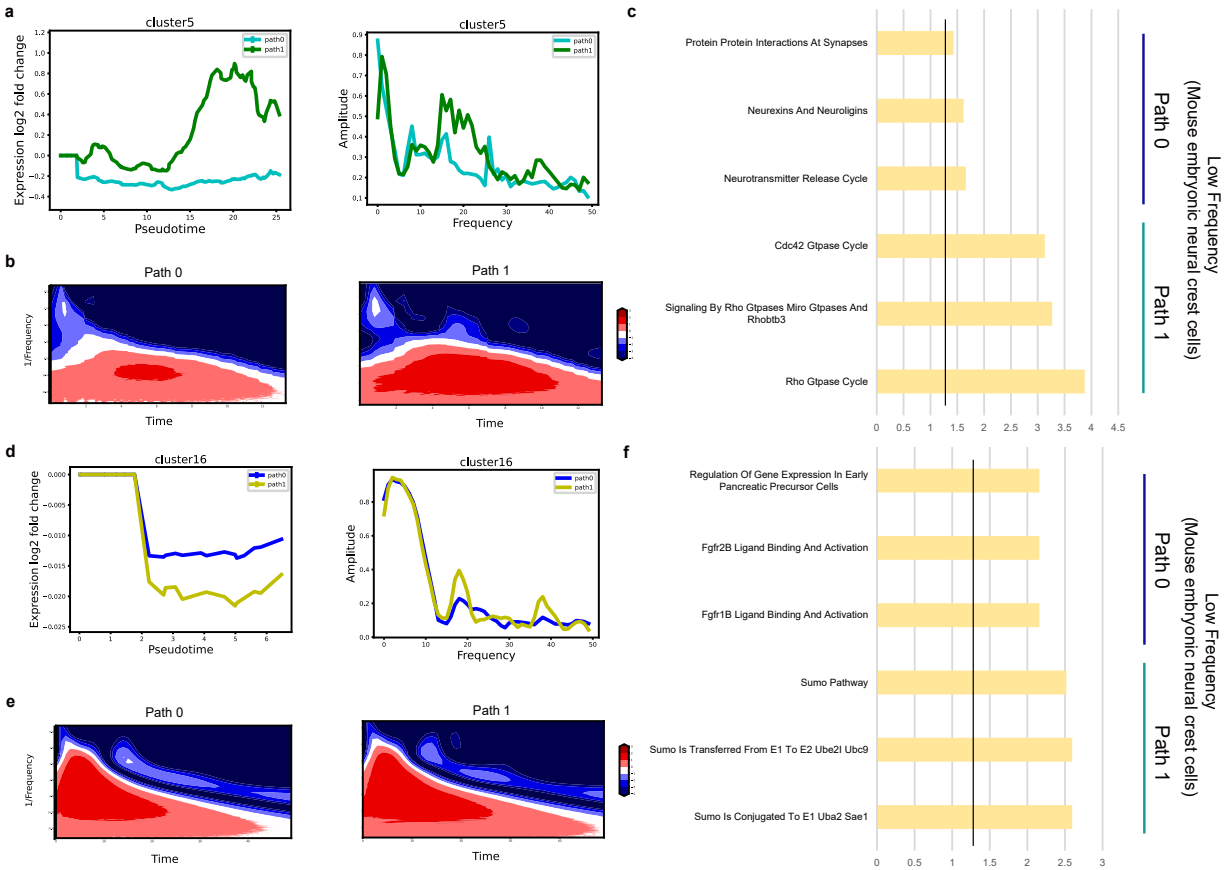
Additional Information

GeneRhythm uncovers gene expression rhythms that shape the dynamic transcriptional landscape in single-cell omics

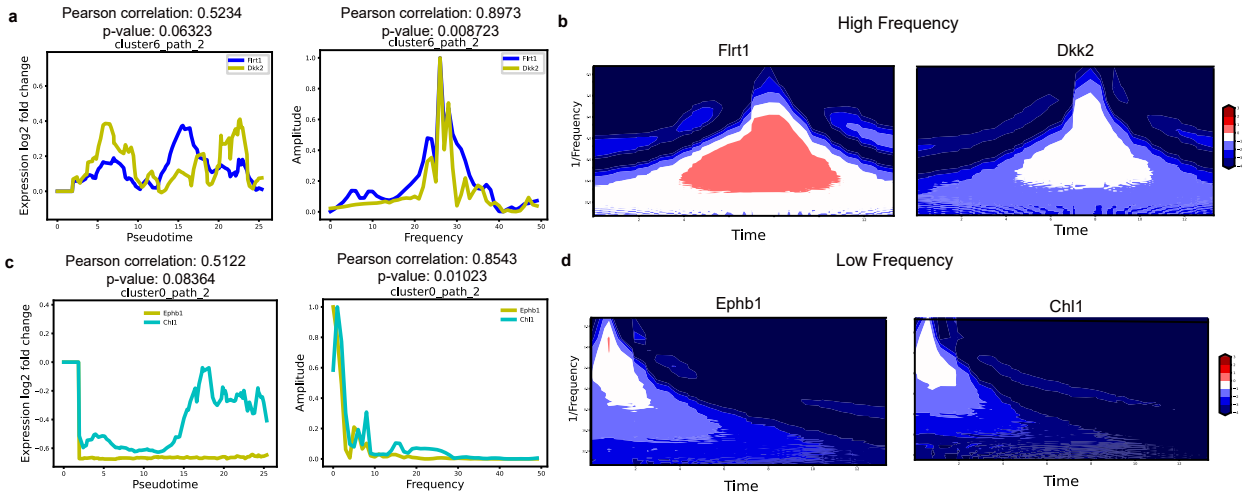
Yang et al.



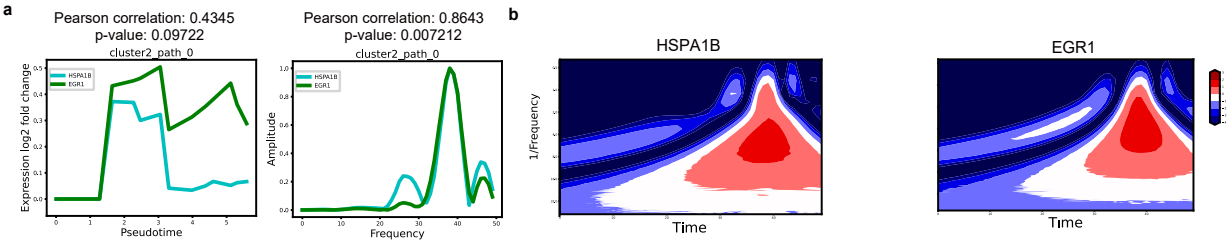
Supplementary Fig. S1: **Rhythm based gene cluster bio-significant analysis on Mouse embryonic blood cells cluster 2 and 9. a-c,** High frequency Mouse embryonic blood cells cluster 2 analysis. Separately, time, overall frequency, wavelet signal and pathway which shows great relation of high frequency bio-activities like immune stimulation. **d-f,** Similar high frequency Mouse embryonic blood cells cluster 9 analysis.



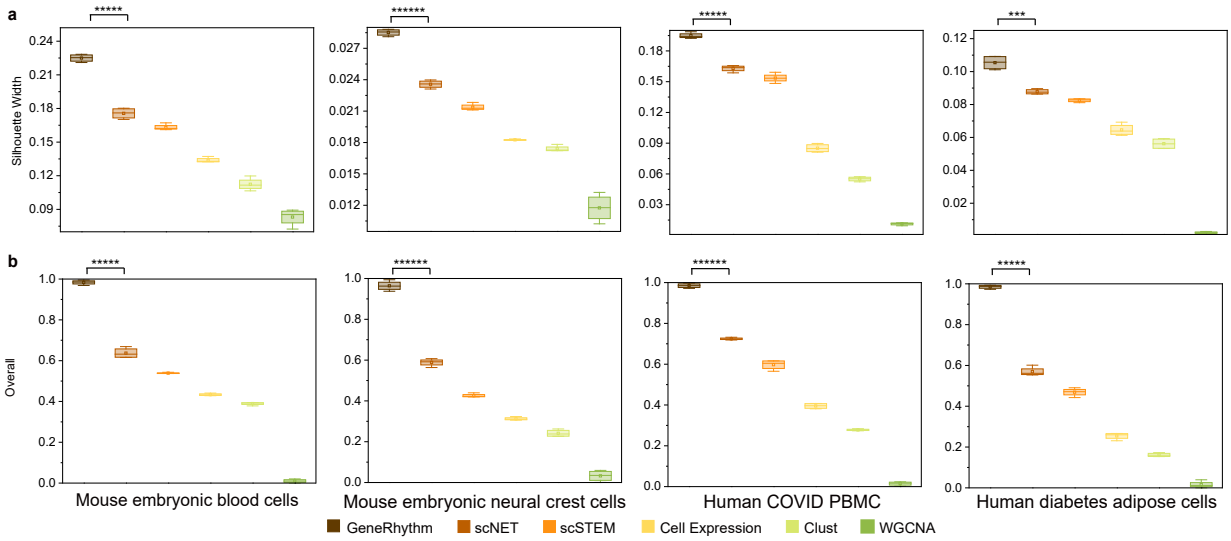
Supplementary Fig. S2: **Rhythm based gene cluster bio-significant analysis on Mouse embryonic neural crest cells cluster 5 and 16. a-c,** Low frequency cluster Mouse embryonic neural crest cells cluster 5 analysis. Separately, time, overall frequency, wavelet signal, pathway which shows great relation of neuron and low frequency bio-activities like metabolism. **d-f,** Similar low frequency Mouse embryonic neural crest cells cluster 16 analysis.



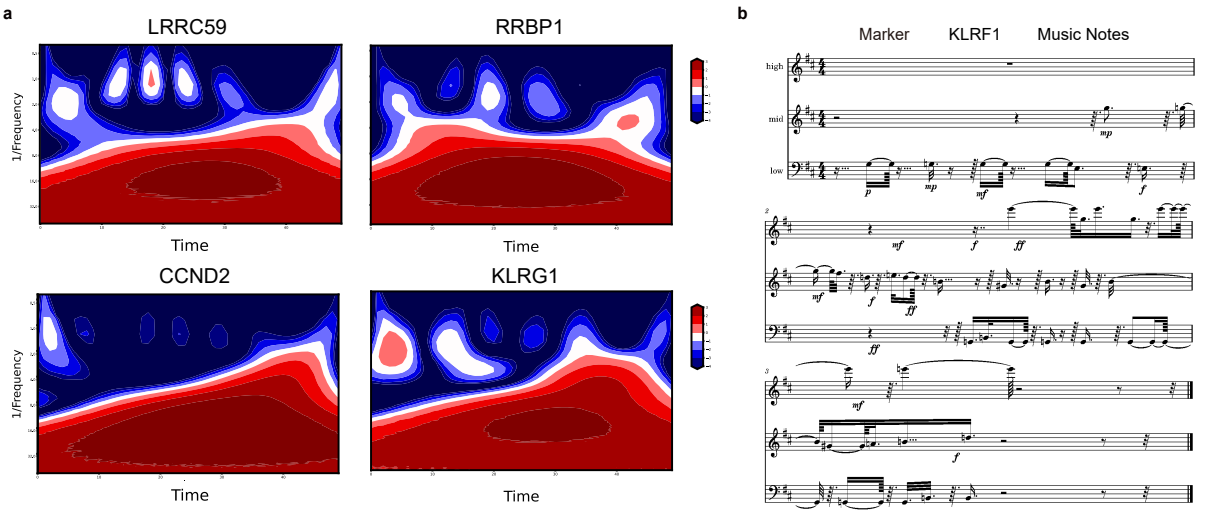
Supplementary Fig. S3: **Genes with different frequency and time information can be clustered together with bio-significance of mouse neuron dataset. a,** Left panel shows the similarity of high frequency genes *Flrt1* and *Dkk2* on time domain information, the pearson correlation is 0.5234 with p-value of 0.06323. Right panel shows the similarity of high frequency genes *Flrt1* and *Dkk2* on overall frequency domain information, the pearson correlation is 0.8973 with p-value of 0.008823. **b,** rhythm patterns of high frequency genes *Flrt1* and *Dkk2*. **c,** Left panel shows the similarity of low frequency genes *Ephb1* and *Chl1* on time domain information, the pearson correlation is 0.5122 with p-value of 0.08364. Right panel shows the similarity of low frequency genes *Ephb1* and *Chl1* on overall frequency domain information, the pearson correlation is 0.8543 with p-value of 0.05023. **d,** rhythm patterns of low frequency genes *Ephb1* and *Chl1*.



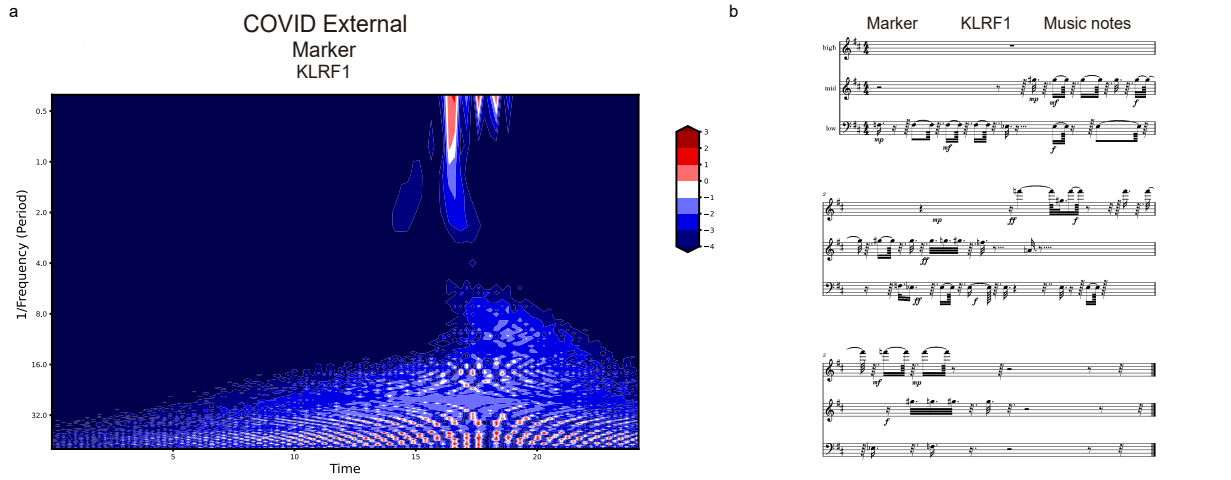
Supplementary Fig. S4: **Genes with different frequency and time information can be clustered with bio-significance on COVID data. a,** Left panel shows the similarity of high frequency genes *HSPA1B* and *EGR1* on time information. Right panel shows the similarity of high frequency genes *HSPA1B* and *EGR1* on overall frequency information. **b,** wavelet signal of high frequency genes *HSPA1B* and *EGR1*.



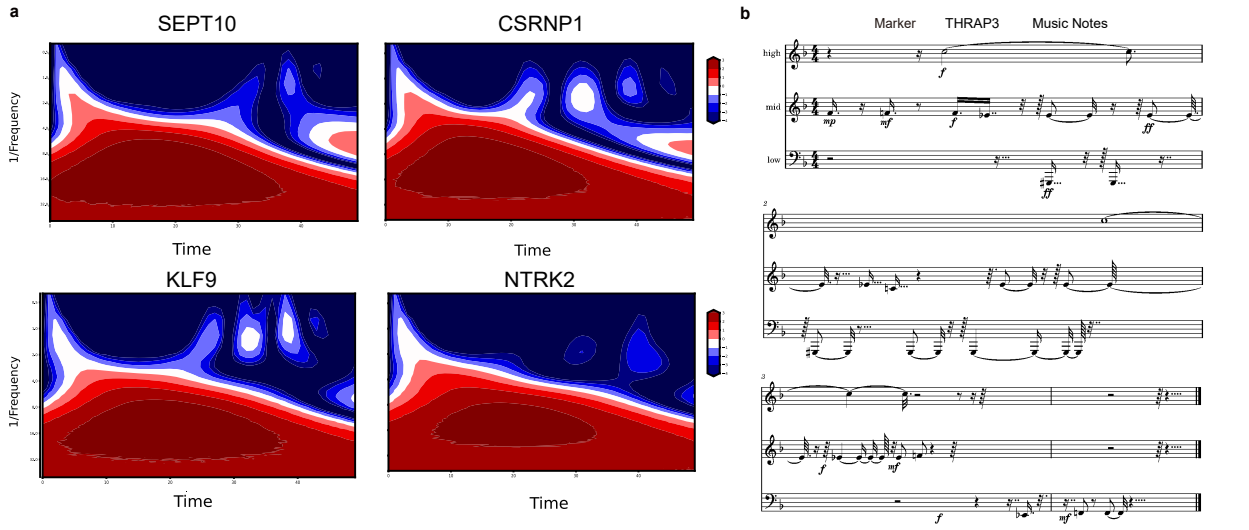
Supplementary Fig. S5: **Benchmarking GeneRhythm for clustering genes with shared expression dynamics. a,** Silhouette Width. **b,** Normalized overall score integrating all evaluation metrics. Benchmarks were performed on mouse embryonic blood cells, mouse embryonic neural crest cells, human COVID PBMCs, and human diabetes adipose cells. Across all datasets and evaluation criteria, GeneRhythm consistently produces gene clusters with higher internal coherence, stronger inter-cluster separation, and greater biological relevance than scNET, scSTEM, cell expression-based clustering, Clust, and WGCNA, reflecting its ability to group genes by shared dynamic expression patterns. Statistical significance was assessed using Welch one-sided t-test. The boxes represent the IQRs, and the solid lines indicate the medians. The whiskers extend to points within 1.5 IQRs of the lower and upper quartiles. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$, ***** $P \leq 0.00001$, ****** $P \leq 0.000001$; ns, not significant.



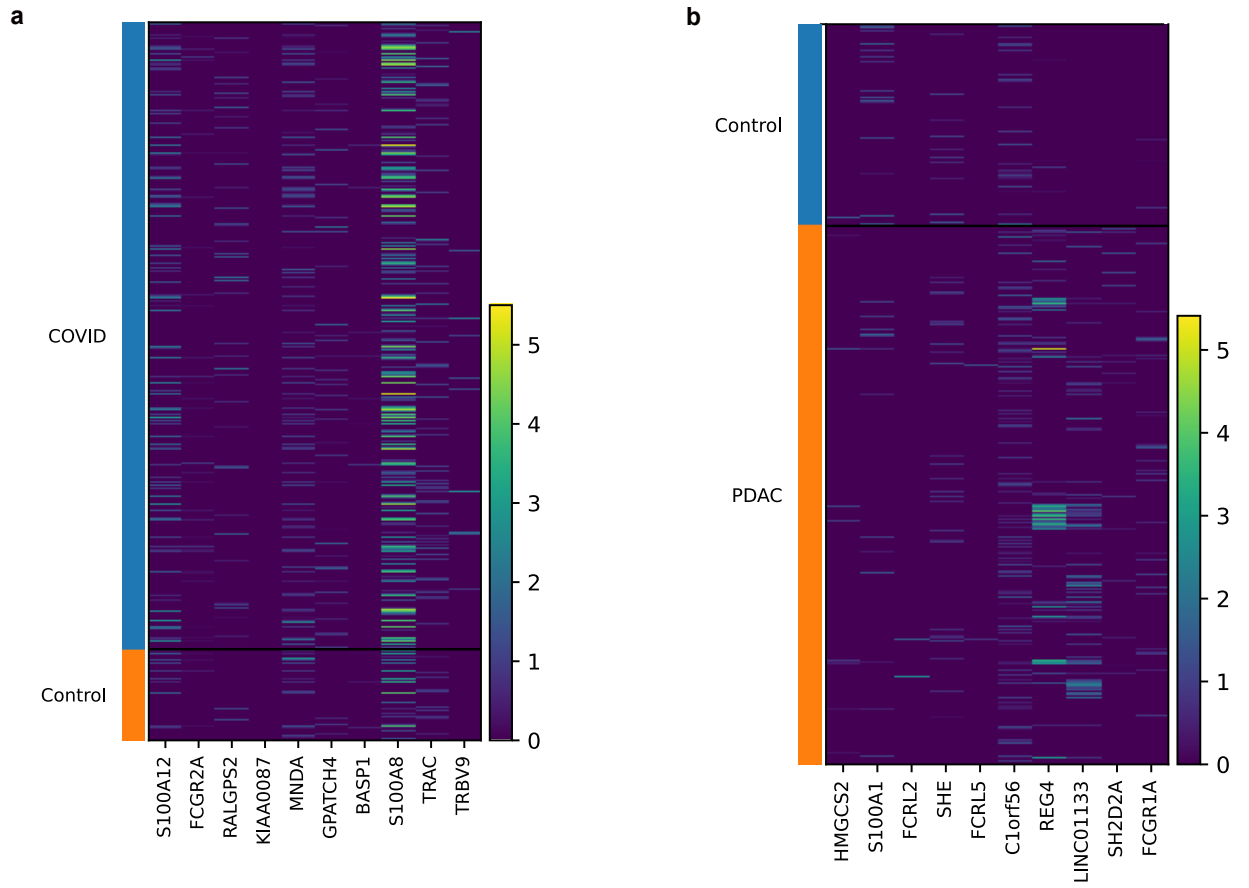
Supplementary Fig. S6: **Rhythm based new bio-marker discovery method analysis of COVID data. a,** Top 4 correlated genes with the COVID. **b,** Music notation of COVID Marker gene *KLRF1* (Full).



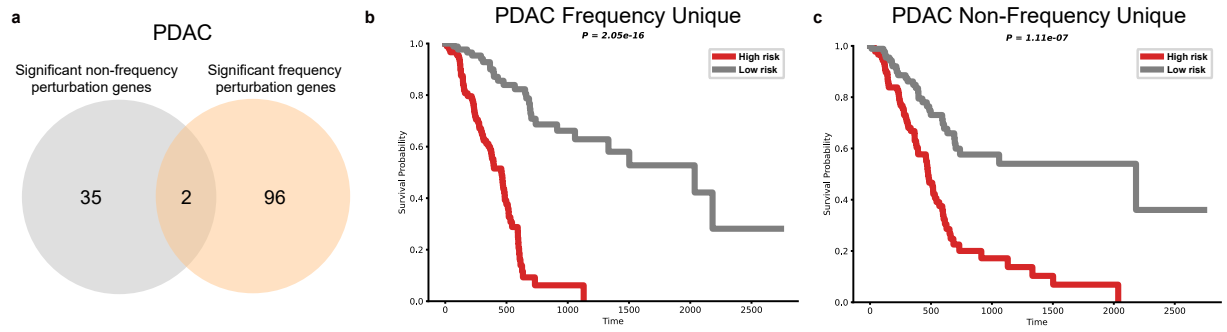
Supplementary Fig. S7: **External differential rhythm pattern and music for COVID-19. a,** Differential rhythm pattern for Marker gene *KLRF1* of external COVID-19 dataset. **b,** Musical notation for Marker gene *KLRF1* of external COVID-19 dataset.



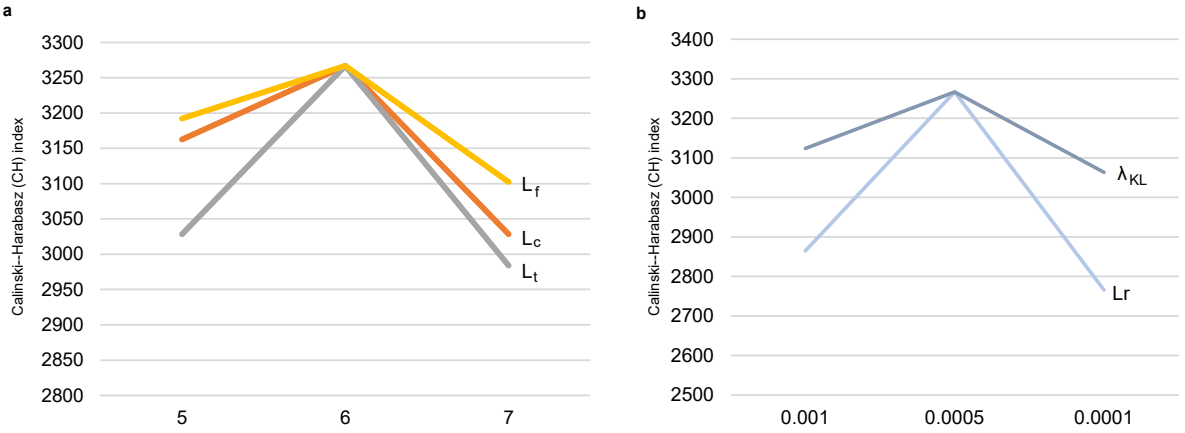
Supplementary Fig. S8: **Rhythm based new bio-marker discovery method analysis of diabetes. a,** Top 4 correlated genes with the diabetes. **b,** Musical notation of Diabetes Marker gene *THRAP3* (Full).



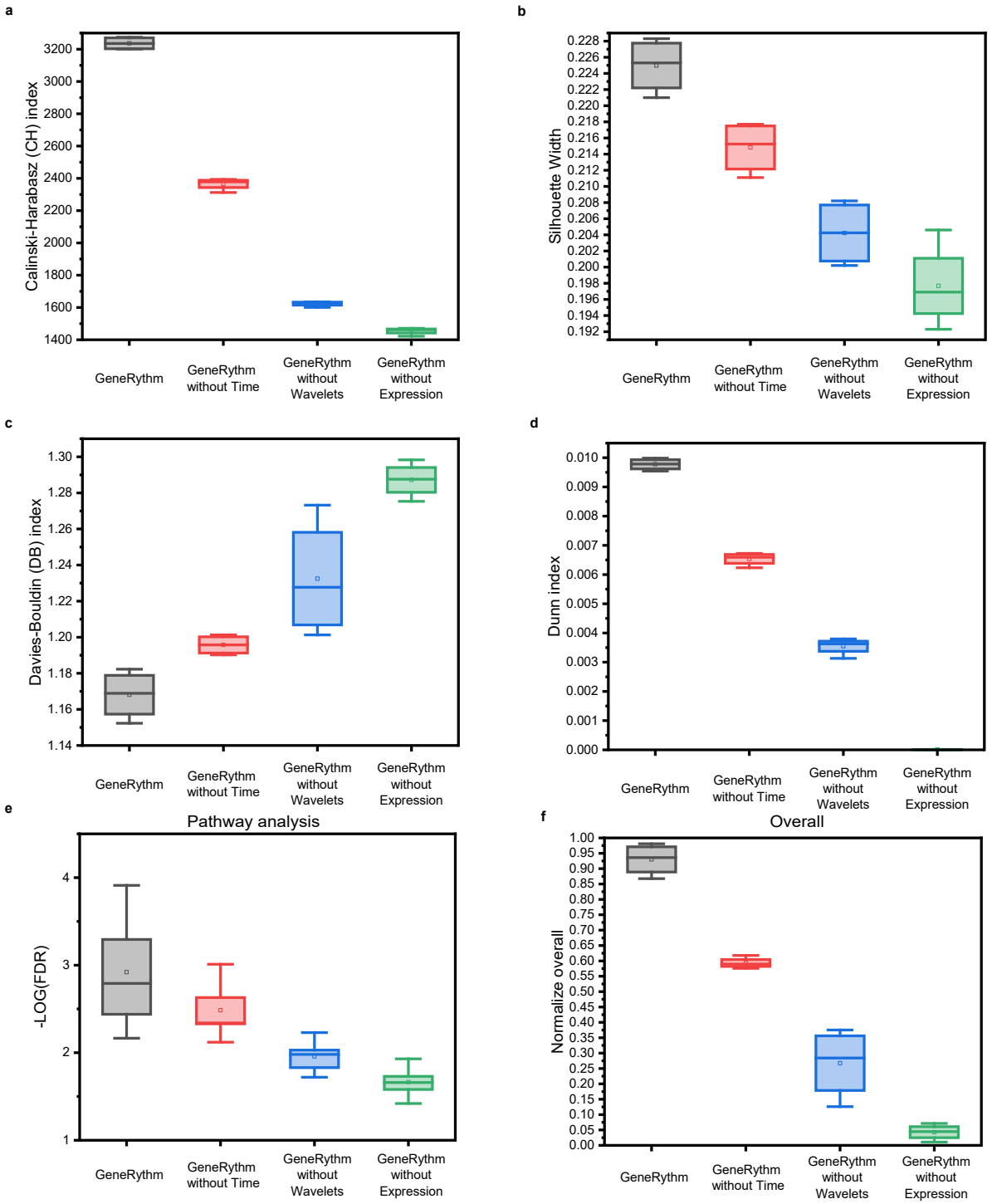
Supplementary Fig. S9: **Gene expression patterns of frequency-based perturbation gene markers. a,** Gene expression of the top frequency-based perturbation gene markers in COVID data. **b,** Gene expression of the top frequency-based perturbation gene markers in PDAC data.



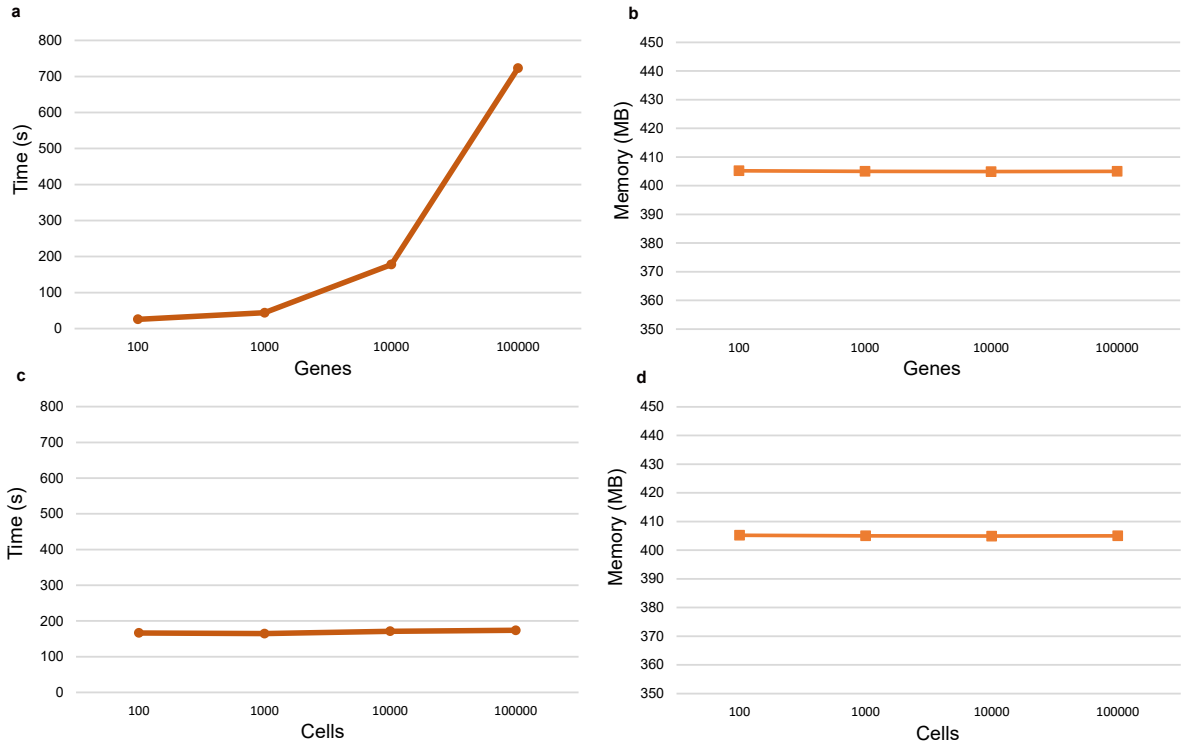
Supplementary Fig. S10: **Frequency-perturbation-identified genes show stronger prognostic stratification than non-frequency genes in PDAC. a,** Venn diagram comparing significant genes identified by non-frequency perturbation and frequency perturbation in PDAC. Frequency perturbation identified 96 unique significant genes, whereas non-frequency perturbation identified 35 unique significant genes, with only 2 genes shared between the two sets. **b,c,** Kaplan–Meier survival analysis of PDAC patients using Cox risk models constructed from genes uniquely identified by frequency perturbation (**b**) or non-frequency perturbation (**c**). The frequency-unique gene set produced stronger survival separation than the non-frequency-unique gene set, with a more significant log-rank association ($P = 2.05 \times 10^{-16}$ versus 1.11×10^{-7}).



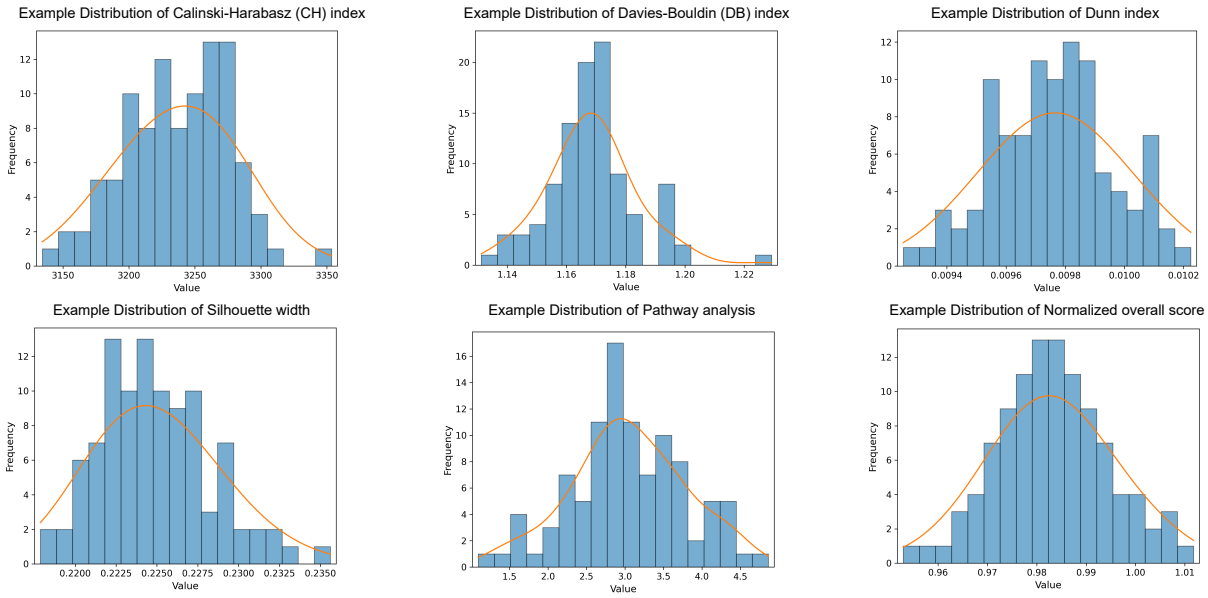
Supplementary Fig. S11: **Hyperparameter sensitivity analysis of GeneRhythm in mouse embryonic blood cells. a,** Clustering performance, measured by the Calinski–Harabasz (CH) index, under different branch-specific latent dimensionalities for the cell-expression (L_c), time-domain (L_t), and frequency-domain (L_f) branches. In all three cases, performance peaked at a latent dimensionality of 6. **b,** Sensitivity of clustering performance to the learning rate (L_r) and KL regularization weight (λ_{KL}), evaluated at 10^{-3}, 5×10^{-4}, and 10^{-4}. Both parameters achieved the best performance at 5×10^{-4}. All experiments were performed on mouse embryonic blood cell data and evaluated using the CH index.



Supplementary Fig. S12: **GeneRhythm ablation study. a-f,** A comprehensive comparative analysis on mouse embryonic blood cells showcases the better performance of GeneRhythm against other ablation situation like GeneRythm without Time, GeneRyth mwithout Wavelets and GeneRythm without Expression. Metrics like RaCalinski-Harabasz (CH) index, Silhouette Width, Davies-Bouldin (DB) index, Dunn index and pathway analysis provide insights into clustering quality on the genes. We have a overall metrics to average all these metrics above.



Supplementary Fig. S13: **GeneRhythm's computation cost. a-b,** We measured the time and memory usage of GeneRhythm on the quantity of genes. Specifically, processing 10,000 genes takes approximately 200 seconds, while processing 100,000 cells takes about 700 seconds. The GPU memory usage is stable about 400 MB. Besides, we measured the time and memory usage of GeneRhythm on the quantity of cells, showing that our method keep stable on time (200 seconds) and GPU memory (400 MB) with the change of cells' amount. These results highlight the scalability of scCross for large datasets. These comparisons were conducted on the Compute Canada platform, utilizing a hardware environment of 1 x NVIDIA P100 Pascal GPU and 2 x Intel E5-2650 v4 Broadwell CPUs.



Supplementary Fig. S14: **Distributions of benchmarking scores.** Examples of benchmarking scores show approximately Gaussian distributions, supporting the use of Welch one-sided t-test in significance analyses presented in benchmarking results.