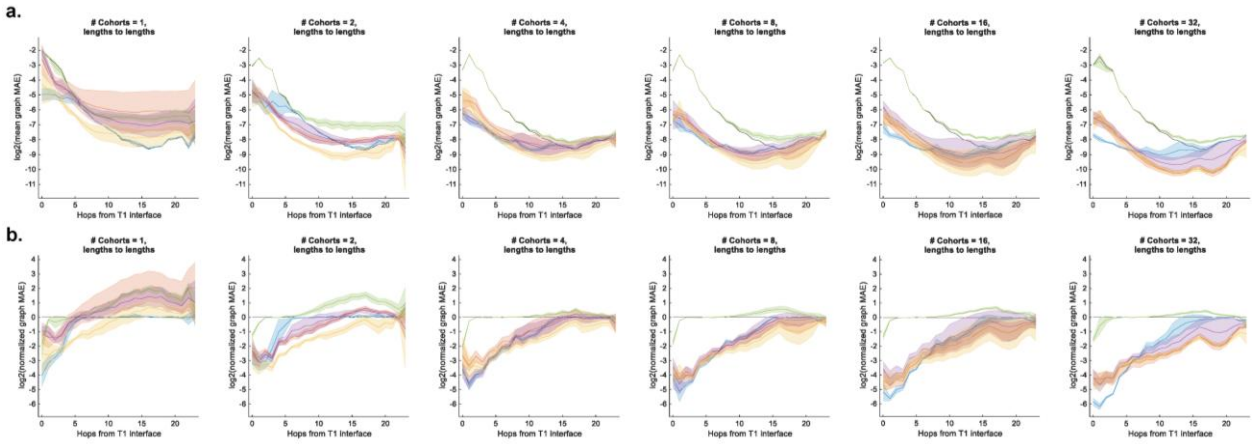
SUPPLEMENTARY INFORMATION

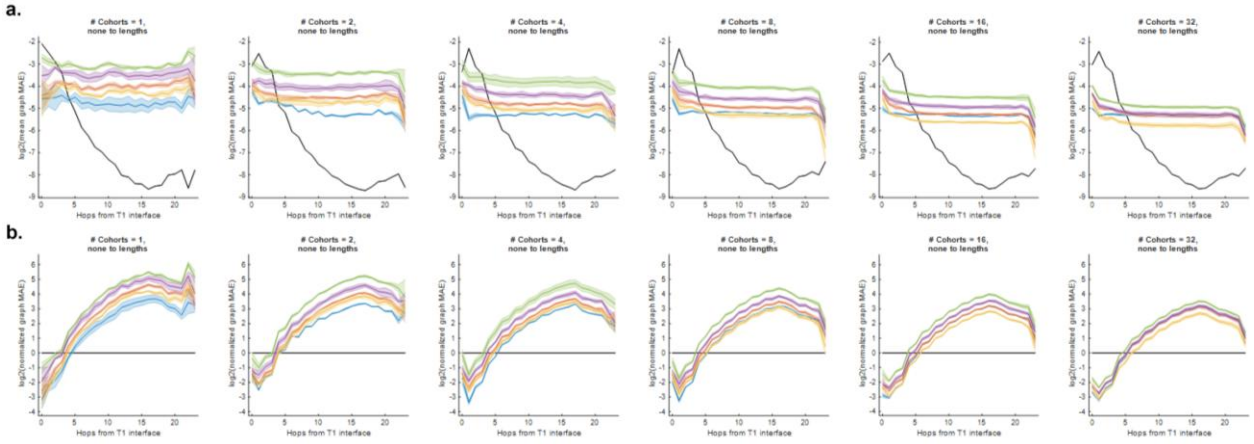
Model	Regime	Learning rate	Hidden	Dropout	Batch
GraphSAGE	Length-informed	0.0006356	128	0	2
GraphSAGE	Length-withheld	0.0018006	128	0.1	2
GAT	Length-informed	0.0005002	64	0	1
GAT	Length-withheld	0.0013725	128	0	4
GIN	Length-informed	0.0003460	128	0	2
GIN	Length-withheld	0.0005578	128	0	2
PNA	Length-informed	0.0004722	64	0	1
PNA	Length-withheld	0.0005104	128	0.1	2

Model	Regime	Learning rate	Scheduler factor	Gradient clipping	Batch
PPGN	Length-informed	0.0005676	0.6	0.1	8
PPGN	Length-withheld	0.0001571	0.3	0.01	4

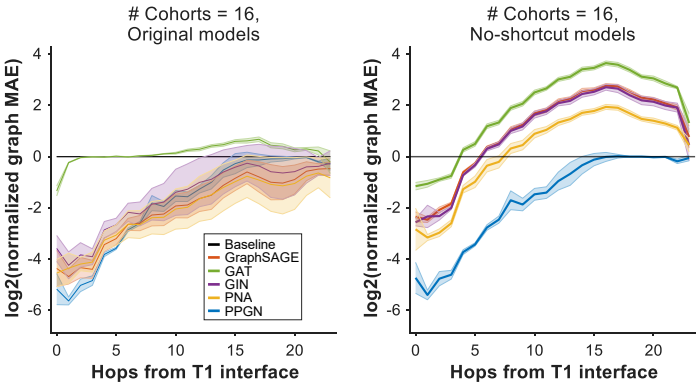
Supplementary Table S1. Optimal hyperparameter configurations. (A) Selected hyperparameters for the four MPNN architectures. (B) Selected hyperparameters for PPGN, which used a different hyperparameter space because it does not use hidden-channel width or dropout in the same way as the MPNN implementations. Hyperparameters were optimized on the original 16-cohort reference dataset, separately for each architecture and input regime, using MATLAB *bayesopt*. For MPNNs, optimized parameters were learning rate, hidden-channel width, dropout, and batch size; network depth was fixed at 16 layers, scheduler_factor at 0.75, and weight_decay at 0. For PPGN, optimized parameters were learning rate, batch size, scheduler factor, and gradient clipping; block_features was fixed at [400, 400, 400] and depth_of_mlp at 2. All downstream analyses, including hexagonality, area elasticity, shear, tissue size, and two-T1 analyses, used the same hyperparameters as the corresponding pre-T1 edge-length-included configuration. “Length-informed” indicates models trained with pre-T1 edge lengths included as input features; “Length-withheld” indicates models trained with pre-T1 edge lengths withheld.



Supplementary Figure S1: Error vs. hop distance at different dataset sizes for edge-length informed models. Prediction error for five GNN architectures and the identity baseline, shown as a function of topological distance (edge hop) from the T1 edge. Each column represents a dataset size defined by the number of training cohorts (1, 2, 4, 8, 16, 32; ≈ 41 graphs per cohort). The top panels (a) report log-transformed raw error: for each hop distance h, edge-level absolute errors were first averaged within each graph, graph-level $\mathrm{MAE}_{\mathrm{model,s}}(\mathrm{h};\mathrm{G})$ values were then averaged across test graphs within each seed, and $\log_2$ was applied before averaging across seeds. The bottom panels (b) report identity-normalized error: graph-level $\log_2$ ratios of $\mathrm{MAE}_{\mathrm{model,s}}(\mathrm{h};\mathrm{G})$ to the hop-specific identity denominator $\mathrm{MAE}_{\mathrm{Id}}(\mathrm{h};\mathrm{G_k})$, averaged across test graphs and seeds. Both sets correspond to models trained with pre-T1 edge lengths provided as input features. Curves indicate mean ± s.d. across n = 5 random seeds. Raw $\mathrm{MAE}_{\mathrm{model}}$ error peaks at the T1 (hop 0) and decays with distance.



Supplementary Figure S2: Error vs. hop distance at different dataset sizes for edge-length withheld models. Prediction error is shown for the same five GNN architectures and identity baseline as in Supplementary Fig. S1, but for models trained without pre-T1 edge lengths. The top panels report log-transformed raw error: for each hop distance h, edge-level absolute errors were first averaged within each graph, graph-level $\mathrm{MAE}_{\mathrm{model,s}}(\mathrm{h};\mathrm{G})$ values were then averaged across test graphs within each seed, and $\log_2$ was applied before averaging across seeds. The bottom panels report identity-normalized error: graph-level $\log_2$ ratios of $\mathrm{MAE}_{\mathrm{model,s}}(\mathrm{h};\mathrm{G})$ to the hop-specific identity denominator $\mathrm{MAE}_{\mathrm{Id}}(\mathrm{h};\mathrm{G_k})$, averaged across test graphs and seeds. Each column corresponds to a dataset size determined by the number of training cohorts (1–32; ≈ 41 graphs per cohort). Curves represent mean ± s.d. across n = 5 random seeds. Without edge-length features, overall accuracy decreases and the effective receptive field (distance over which $\mathrm{nMAE}_{\mathrm{Id}} < 0$) shortens, indicating that explicit geometric cues substantially enhance long-range predictive power.



Supplementary Figure S3: Architecture-specific edge-length shortcut ablations. Hop-resolved identity-normalized prediction error is shown for models trained on the 16-cohort edge-length-informed dataset. The left panel shows the original models; the right panel shows the corresponding ablated models. For GraphSAGE, GAT, GIN, and PNA, raw edge attributes were removed from the final edge-regression head, while the message-passing backbone was otherwise left intact. For PPGN, the first block skip connection was disabled while later block skips remained active. Curves show mean ± s.d. across $n = 5$ random seeds. The MPNN ablations strongly disrupted distal fallback, whereas PPGN retained substantial fallback.