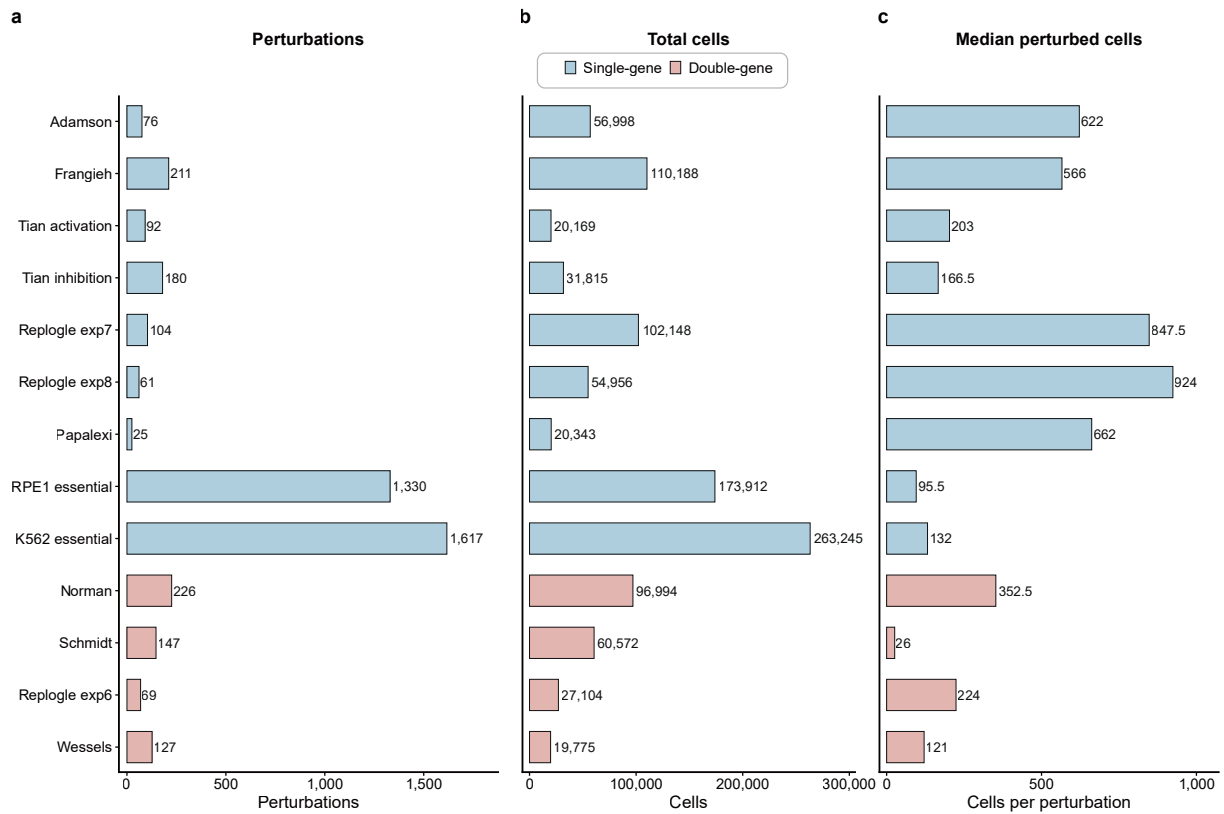
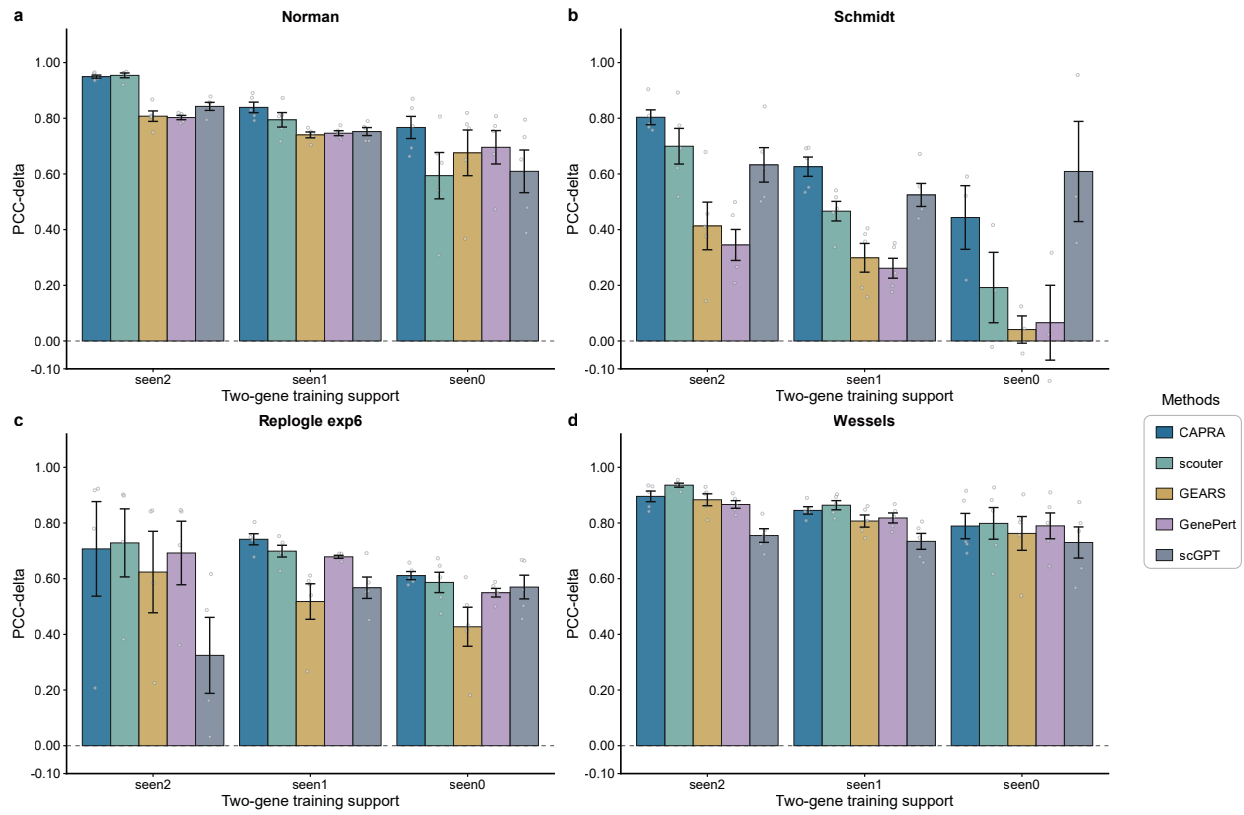




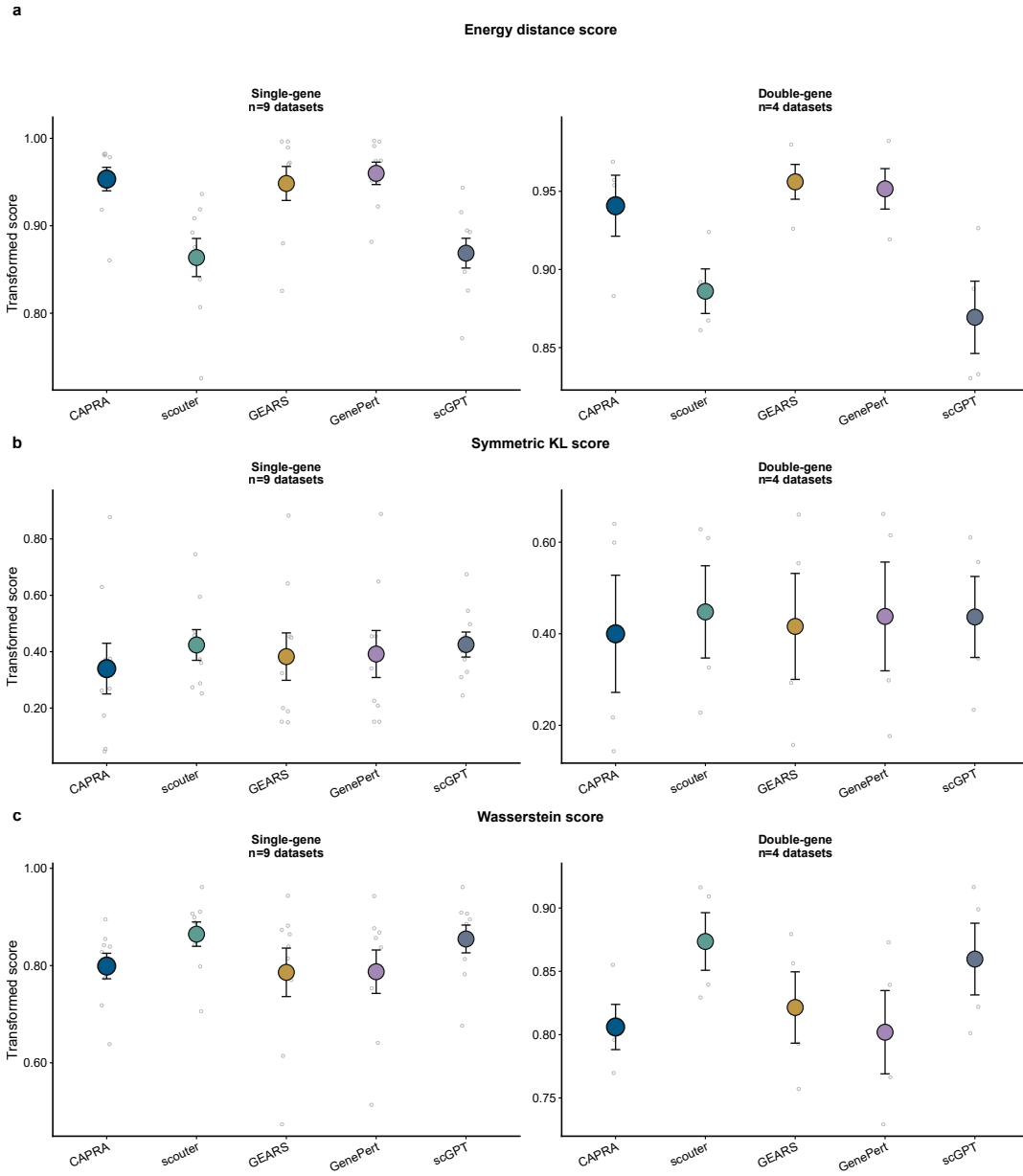
1

2 **Supplementary Fig. S1 | Dataset scale and perturbation-cell support.** **a**, Number of non-control perturbation labels
3 per dataset. **b**, Total cells in the processed expression object, including control and perturbed cells. **c**, Median perturbed cells
4 per perturbation. Rows are genetic-perturbation datasets ($n = 13$); labels report plotted values.



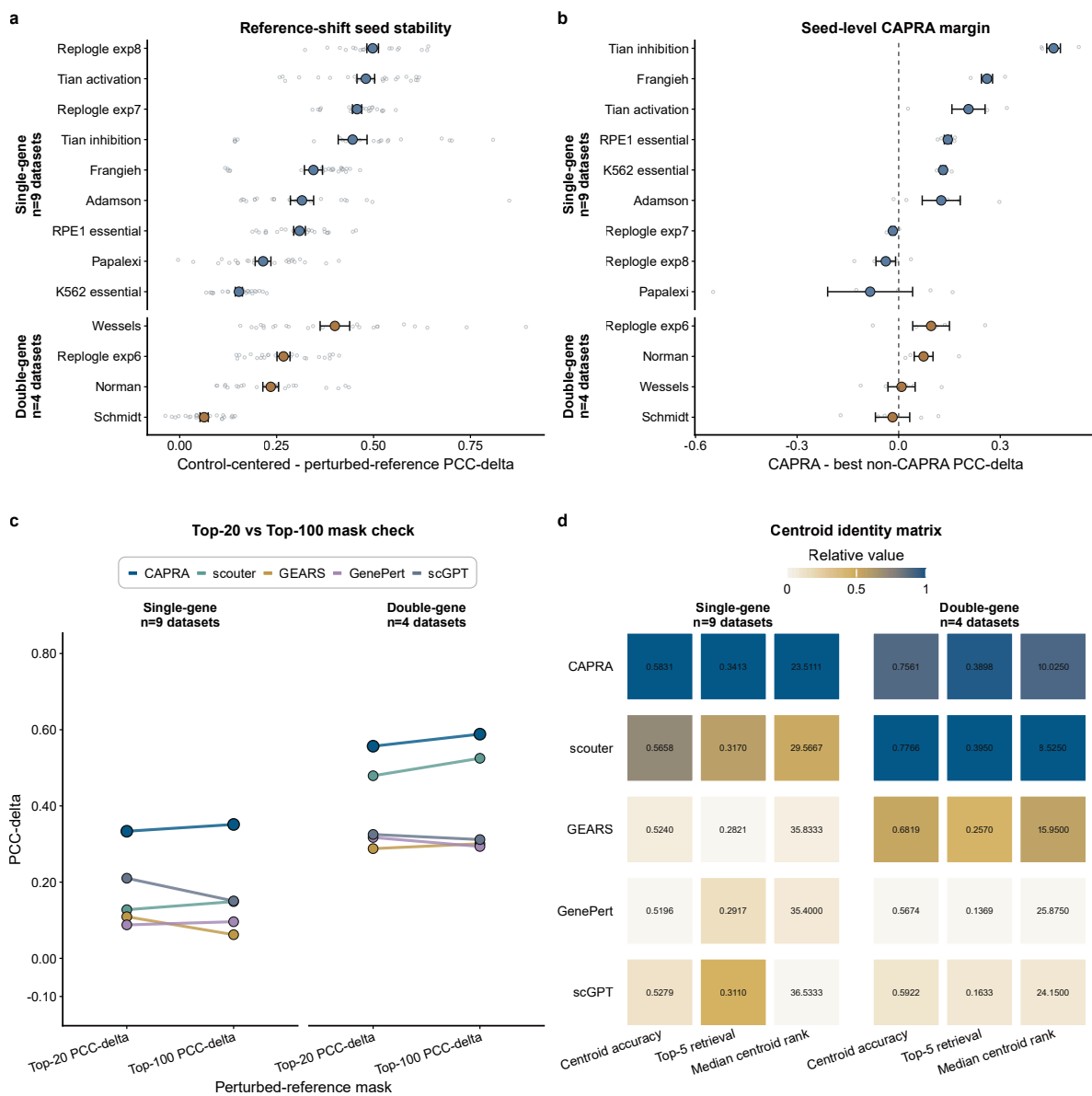
5

6 **Supplementary Fig. S2 | Dataset-specific double-gene support tiers.** a, Norman PCC-delta across seen2, seen1 and
7 seen0 double-gene pairs. b, Schmidt PCC-delta under the same support-tier grouping. c, Replogle exp6 PCC-delta. d, Wessels
8 PCC-delta. Bars show seed-averaged method performance within each group; points show seed values and intervals show mean
9 $\pm$ s.e.m. across seeds.



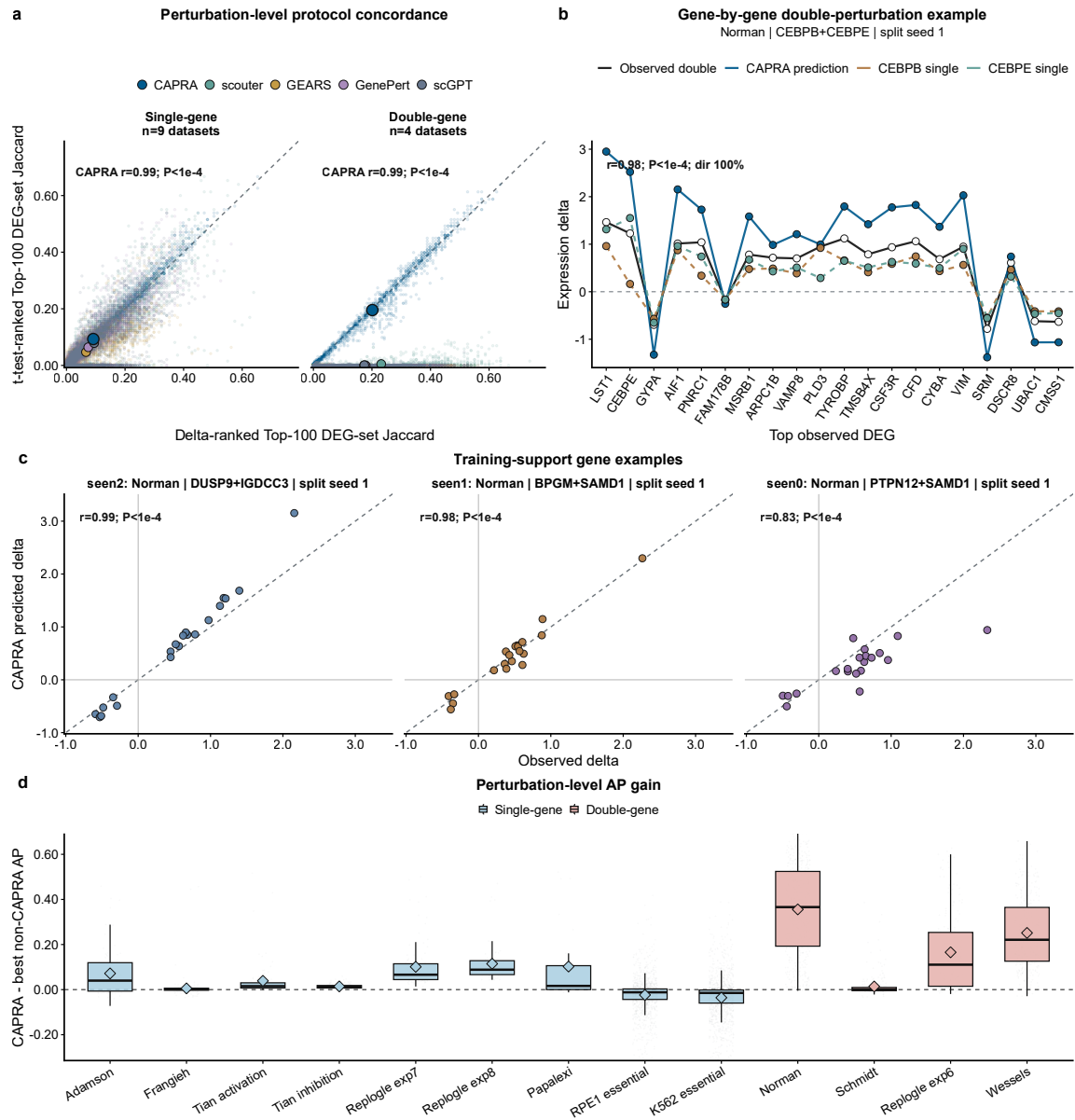
10

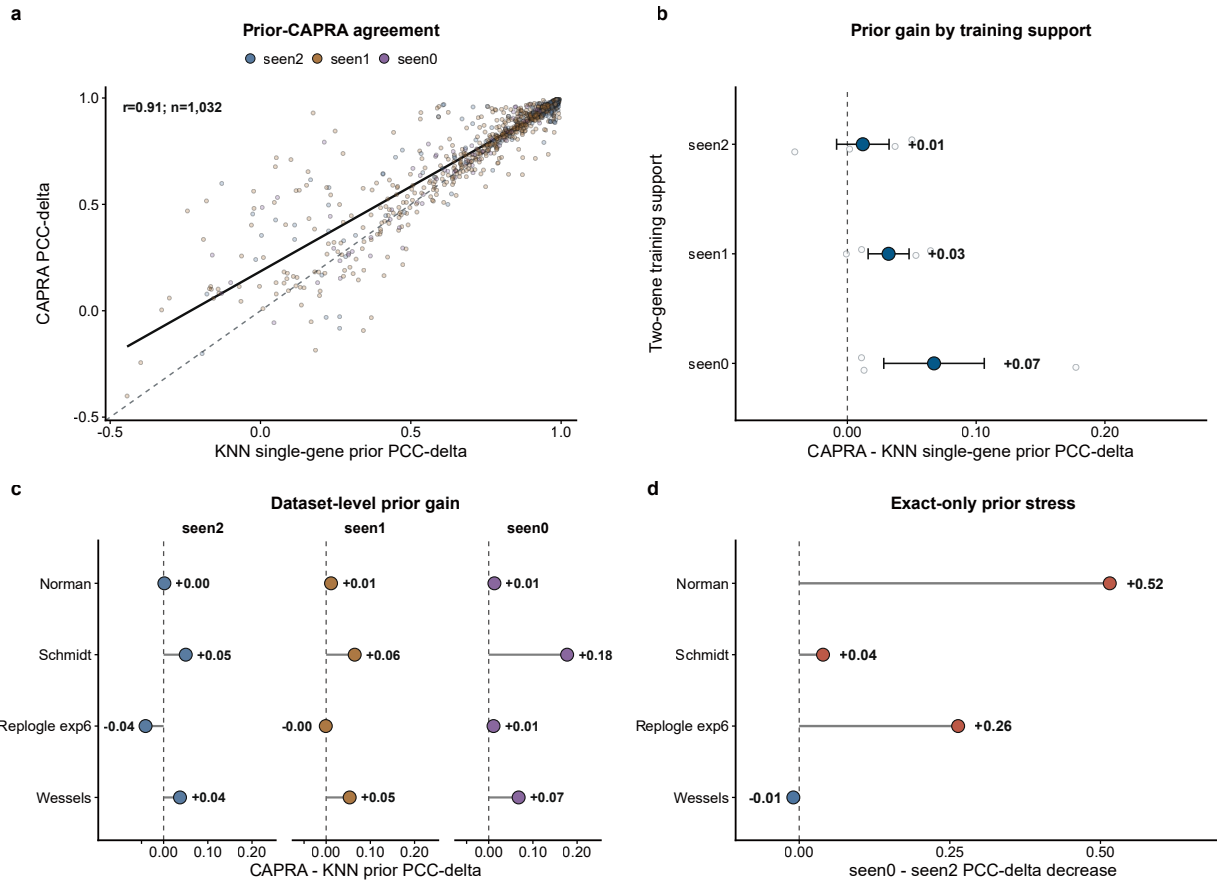
11 **Supplementary Fig. S3 | Additional distributional evaluation checks.** **a**, Energy-distance score after transforming the
 12 lower-is-better raw metric to a 0–1 higher-is-better display scale. **b**, Symmetric KL score under the same matched-denominator
 13 transformation. **c**, Wasserstein score under the same transformation. Grey points show dataset values, coloured points show
 14 mean values across datasets and intervals show mean $\pm$ s.e.m. across datasets.



15

16 **Supplementary Fig. S4 | Perturbed-reference seed-level and centroid readouts.** **a**, Seed-level score shift from
 17 conventional control-centred Top-100 PCC-delta to perturbed-reference Top-100 PCC-delta for each method and dataset. **b**,
 18 Seed-level CAPRA margin relative to the highest non-CAPRA value on the Top-100 perturbed-reference metric. **c**, Comparison
 19 between Top-20 and Top-100 perturbed-reference masks. **d**, Centroid-identity readouts computed from the same perturbed-
 20 reference predictions, including centroid accuracy, top-5 retrieval and median centroid rank.





27

28 **Supplementary Fig. S6 | Anchor-support and exact-only stress readouts.** a, Perturbation-level agreement between
 29 the KNN single-gene-anchor readout and full CAPRA predictions. b, Full-CAPRA minus KNN-anchor difference after grouping
 30 double-gene perturbations by single-gene training support. c, Dataset-level full-CAPRA minus KNN-anchor difference across
 31 seen2, seen1 and seen0 groups. d, Exact-only anchor stress, summarized as seen0-minus-seen2 PCC-delta decrease by dataset.

Supplementary Table 1 | Computational profiling on the Norman dataset.

Method	GPU	Runtime (min)	Peak RSS (GiB)	Peak PSS (GiB)	Peak GPU memory (GiB)	Peak GPU util. (%)
CAPRA	Yes	6.43 ± 0.29	137.53 ± 0.35	19.02 ± 0.07	$2.63 \pm <0.01$	47.00 ± 1.58
Scouter	Yes	12.14 ± 0.94	21.32 ± 0.30	21.60 ± 0.35	5.42 ± 0.10	29.00 ± 9.81
GEARS	Yes	47.20 ± 0.74	23.88 ± 0.12	23.82 ± 0.11	$2.33 \pm <0.01$	65.00 ± 1.34
GenePert	No	0.79 ± 0.04	14.93 ± 0.64	14.89 ± 0.64	N.A.	N.A.
scGPT	Yes	324.01 ± 9.91	23.10 ± 0.02	23.08 ± 0.02	20.73 ± 0.09	100.00 ± 0.34

Values summarize a Norman five-seed runtime and resource-use profiling comparison, not a separate predictive-performance benchmark. Entries with $\pm$ denote mean $\pm$ s.e.m. Runtime is wall-clock time. CPU memory is reported as peak resident set size (RSS) and peak proportional set size (PSS); PSS apportions shared pages and is included because RSS can over-count shared-memory data-loading pages. GPU memory and utilization were measured on an NVIDIA GeForce RTX 3090 with 24 GiB memory. N.A., not applicable.