Supplementary Information

De novo design of neutralizing protein binders against CCL26 and POSTN, two convergent targets in eosinophilic esophagitis

This document contains supplementary figures S1–S12, supplementary tables S1–S5, and supplementary media (two 3D rotation videos). All data products are available as project artifacts.

Supplementary Figures

Figure S1 — Meta-analysis effect-size landscape

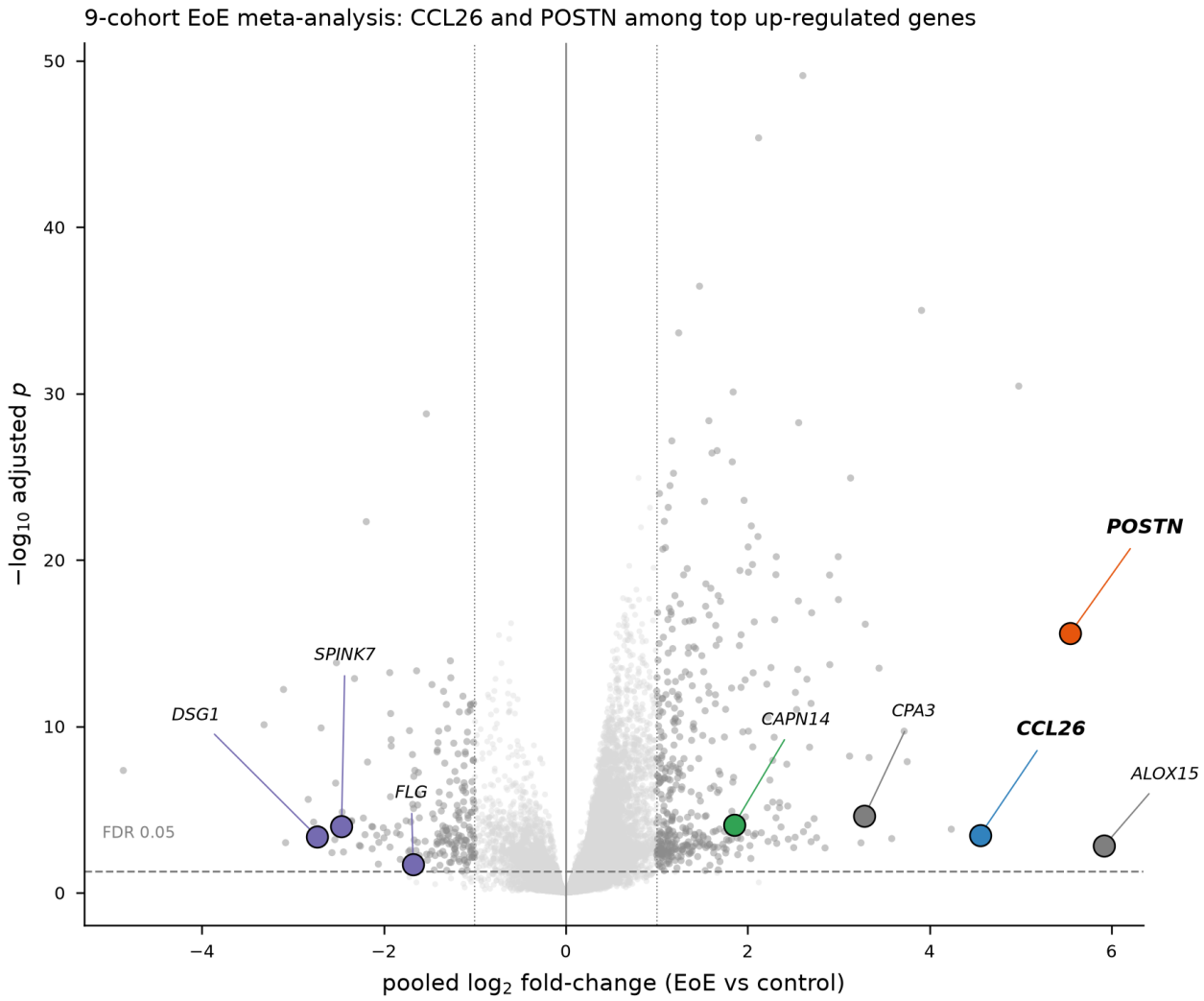

Figure S1. Volcano plot of the nine-cohort EoE random-effects meta-analysis (25,654 genes): pooled $\log_2$ fold-change vs. $-\log_{10}$ adjusted p. CCL26 and POSTN (bold, colored) are among the most strongly up-regulated genes; concordantly down-regulated barrier genes (DSG1, SPINK7, FLG) appear on the left. Dashed horizontal line = FDR 0.05; dotted vertical lines = ±1 $\log_2$FC.

Figure S2 — Literature attention vs. omics effect size

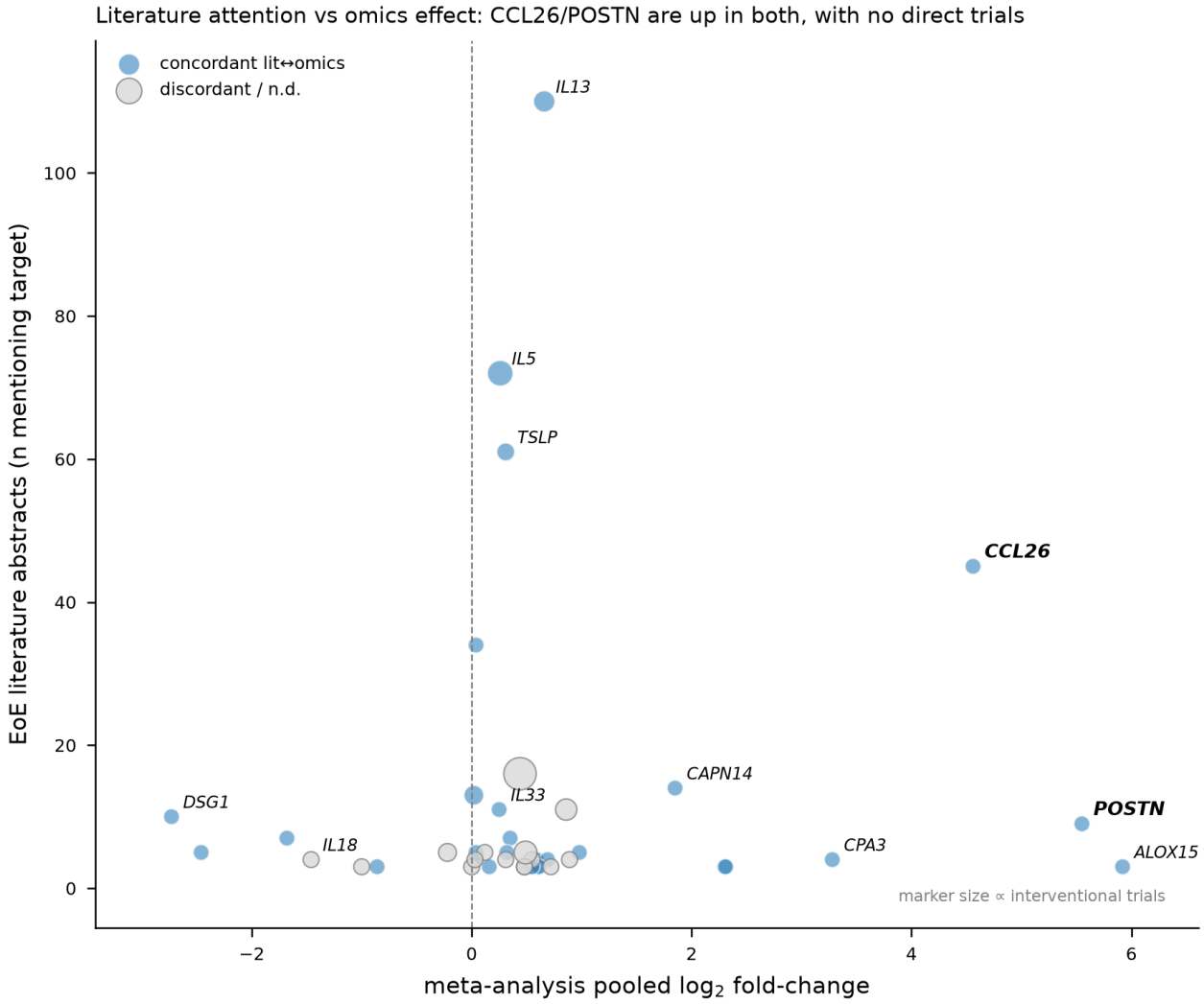

Figure S2. Cross-reference of the meta-analysis against the EoE literature corpus (591 abstracts, 2015–2026) and interventional-trial landscape (157 trials). x-axis: meta-analysis pooled $\log_2$FC; y-axis: number of abstracts mentioning the target; marker size proportional to interventional-trial count; color indicates literature↔omics direction concordance. CCL26 and POSTN combine high omics effect size with zero direct interventional trials, unlike the heavily-studied cytokine nodes (IL-13, IL-5, TSLP).

Figure S3 — Single-cell cell-of-origin

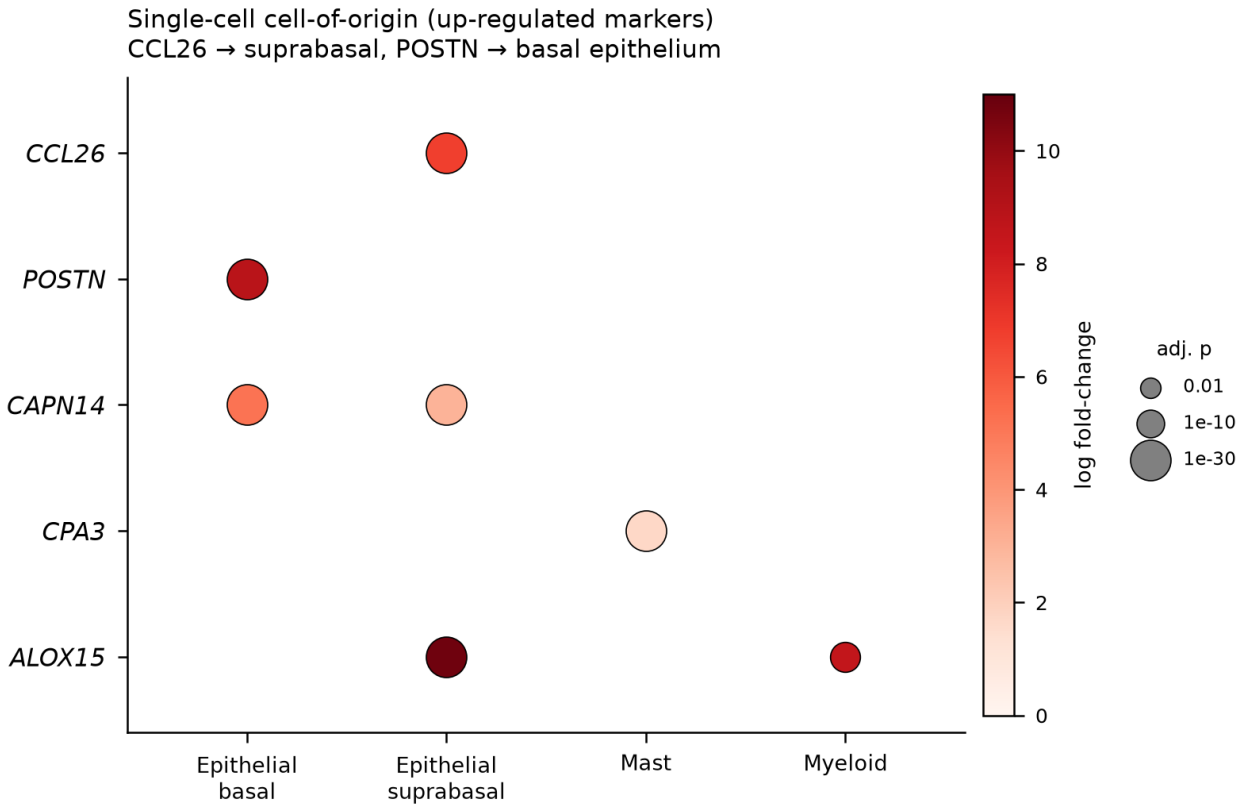

Figure S3. Single-cell differential-expression dot plot for the up-regulated candidate targets across EoE epithelial and effector-cell compartments. Dot color = log fold-change, dot size = $-\log_{10}$ adjusted *p*. CCL26 localizes to suprabasal epithelium and POSTN to basal epithelium; CAPN14 spans both epithelial layers, CPA3 marks mast cells, and ALOX15 marks suprabasal epithelium and myeloid cells. (Only genes with up-regulated single-cell markers in the source table are shown.)

Figure S4 — Omics druggability shortlist

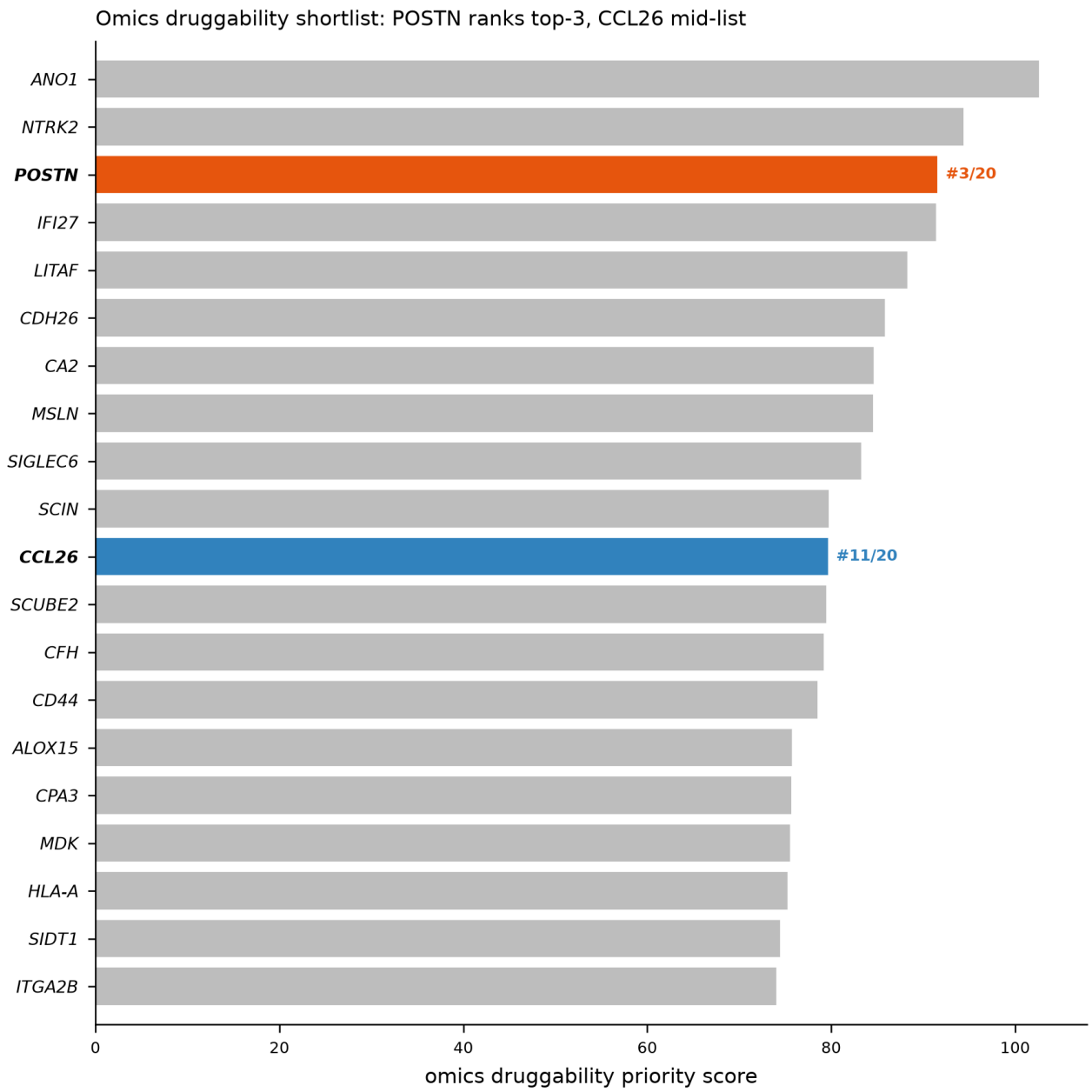

Figure S4. Omics-derived druggability priority score for the 20-gene shortlist (filtered for effect size, direction concordance, single-cell specificity, and antibody tractability). POSTN ranks 3rd of 20 and CCL26 11th of 20; both are secreted and suitable for a ligand-trap modality.

Figure S5 — RFdiffusion backbone quality control

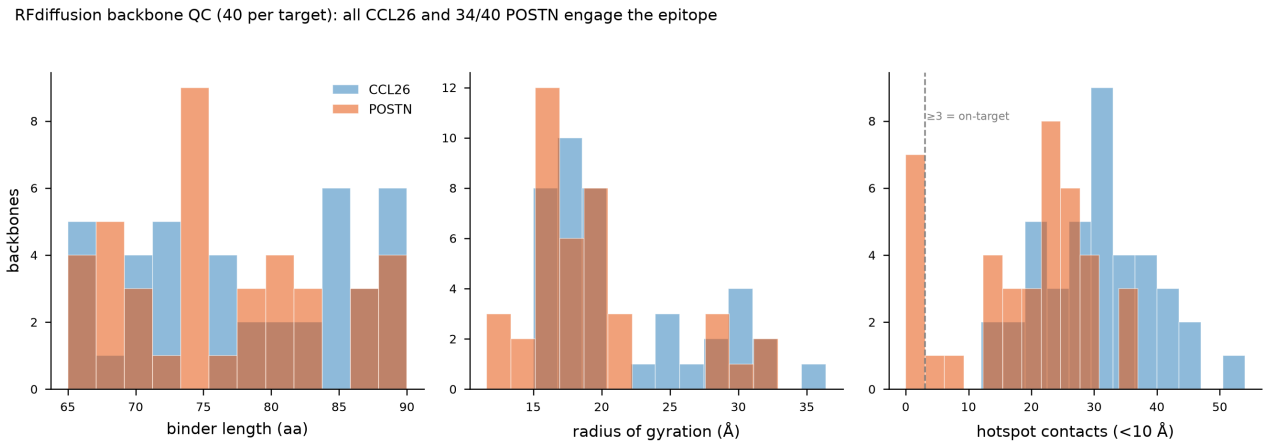

Figure S5. Distributions of binder length, radius of gyration, and hotspot contacts (binder Cα within 10 Å of hotspot Cα) for 40 RFdiffusion backbones per target. All 40 CCL26 and 34/40 POSTN backbones engage ≥3 hotspot residues (dashed line).

Figure S6 — Epitope definition and solvent accessibility

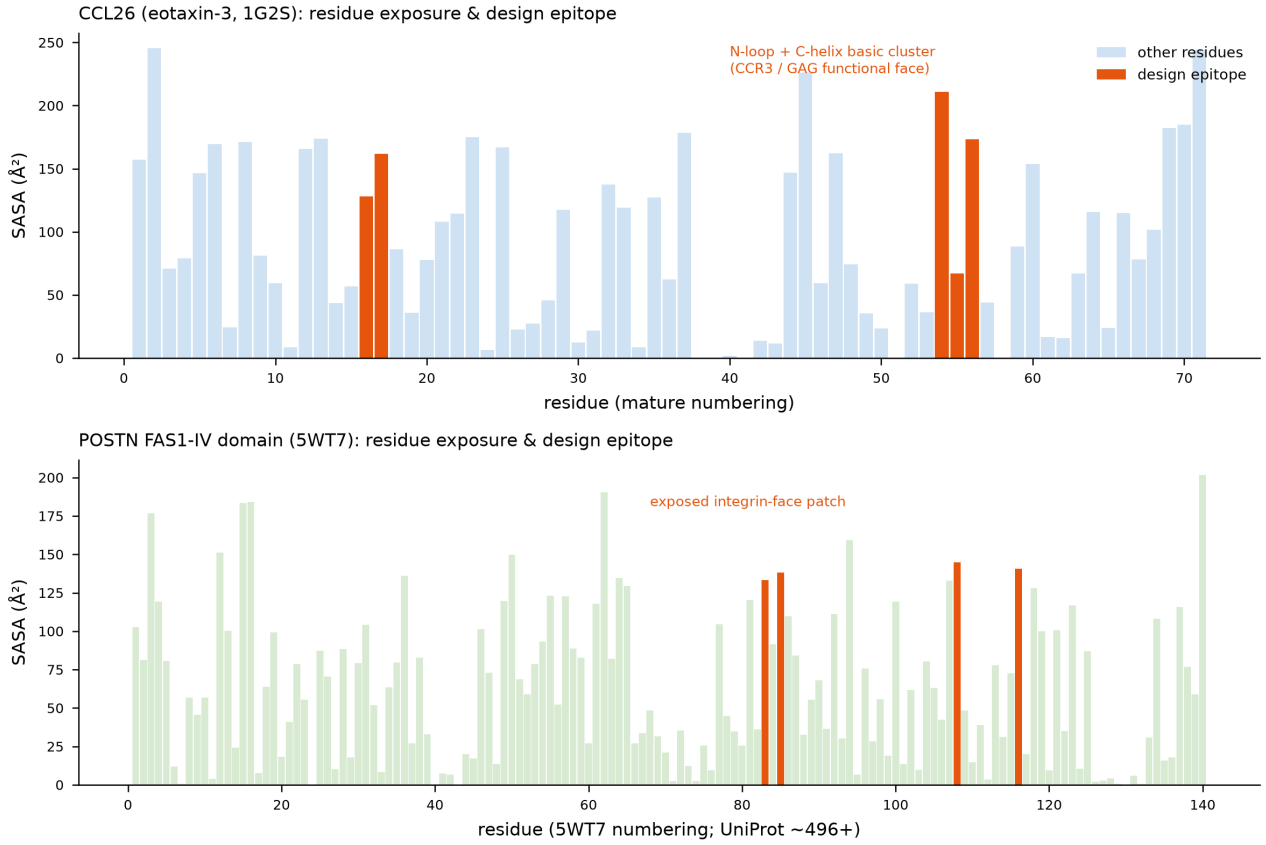

Figure S6. Per-residue solvent-accessible surface area (Shrake–Rupley) for the CCL26 and POSTN target structures, with design epitope hotspots highlighted. CCL26 hotspots (16, 17, 54, 55, 56) span the N-loop and C-terminal α-helix basic cluster (CCR3/GAG face); POSTN hotspots (83, 85, 108, 116) form an exposed patch on the FAS1-IV integrin-interaction surface. All epitope residues are surface-exposed.

Figure S7 — SolubleMPNN score distributions

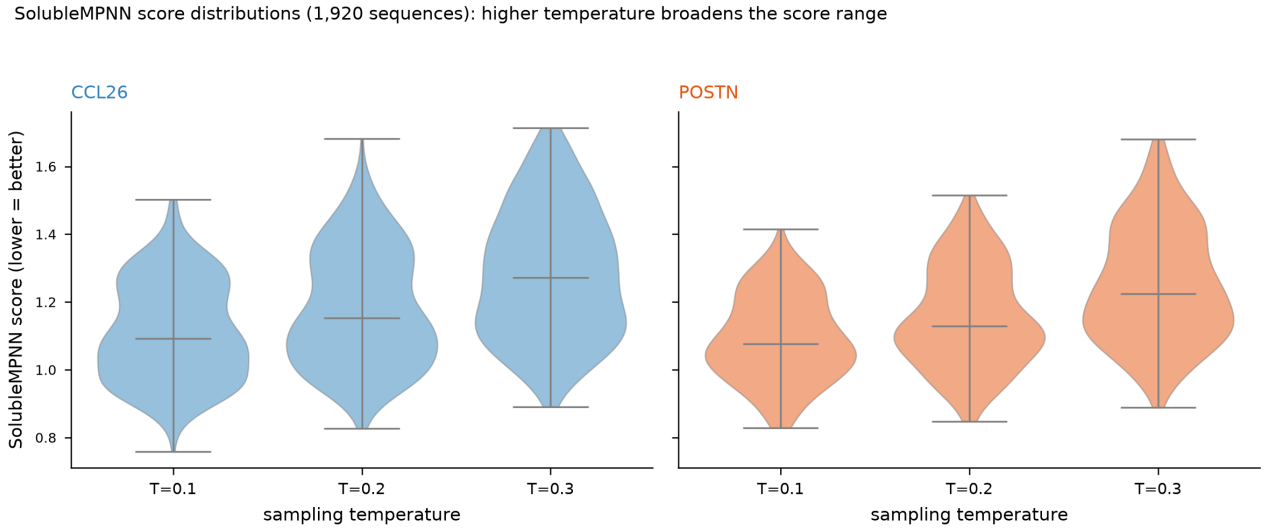

Figure S7. SolubleMPNN score distributions (lower = better) for all 1,920 designed sequences, split by target and sampling temperature (T = 0.1, 0.2, 0.3). Higher temperature broadens the score distribution as expected.

Figure S8 — Sequence-complexity filter

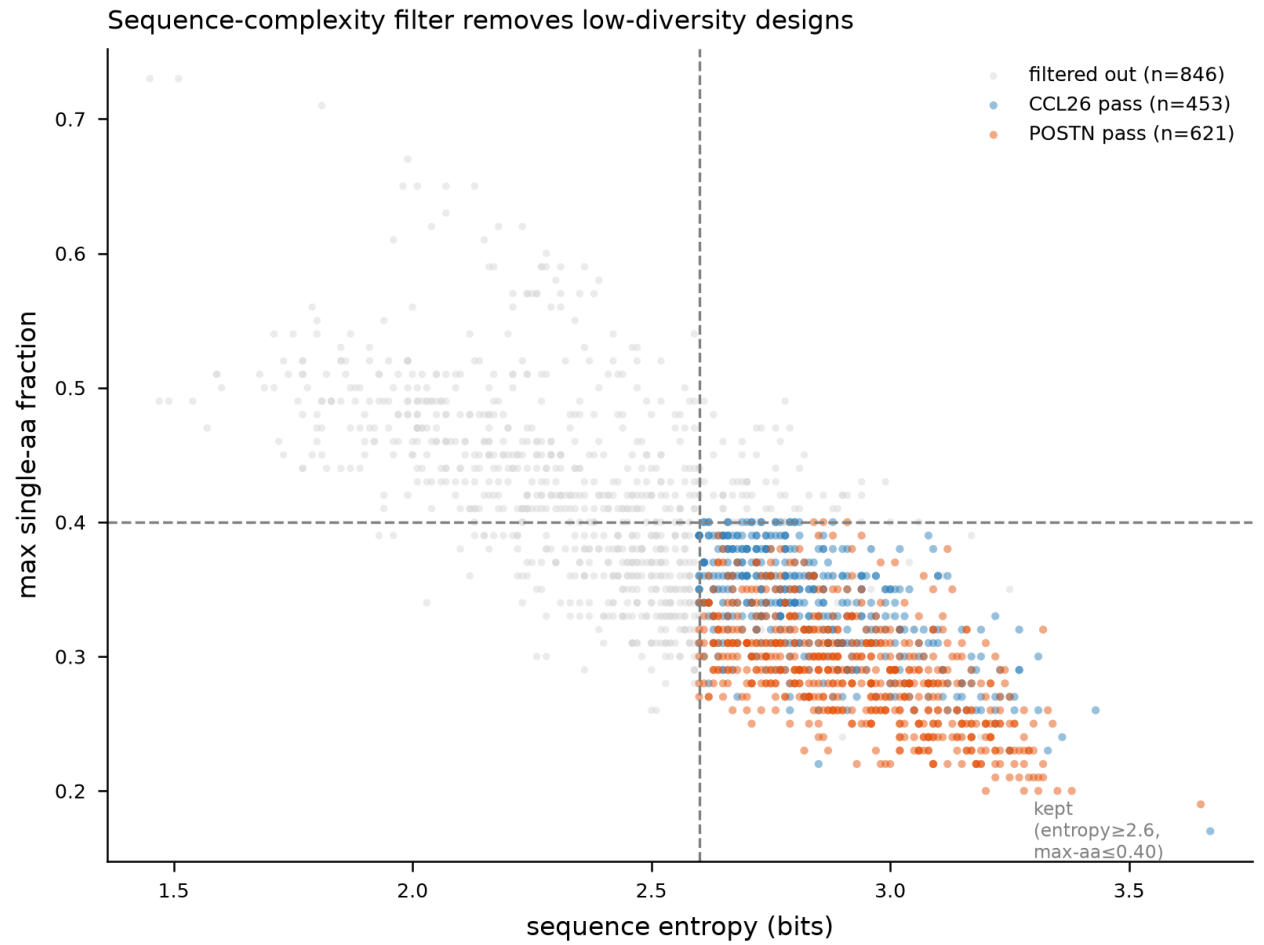

Figure S8. Sequence entropy vs. maximum single-amino-acid fraction for all designed sequences. The complexity filter (entropy ≥ 2.6 bits, max single-aa fraction ≤ 0.40, max single-residue run ≤ 5; dashed lines) removes low-diversity designs, retaining 1,074/1,920 sequences.

Figure S9 — Full fold-back confidence landscape

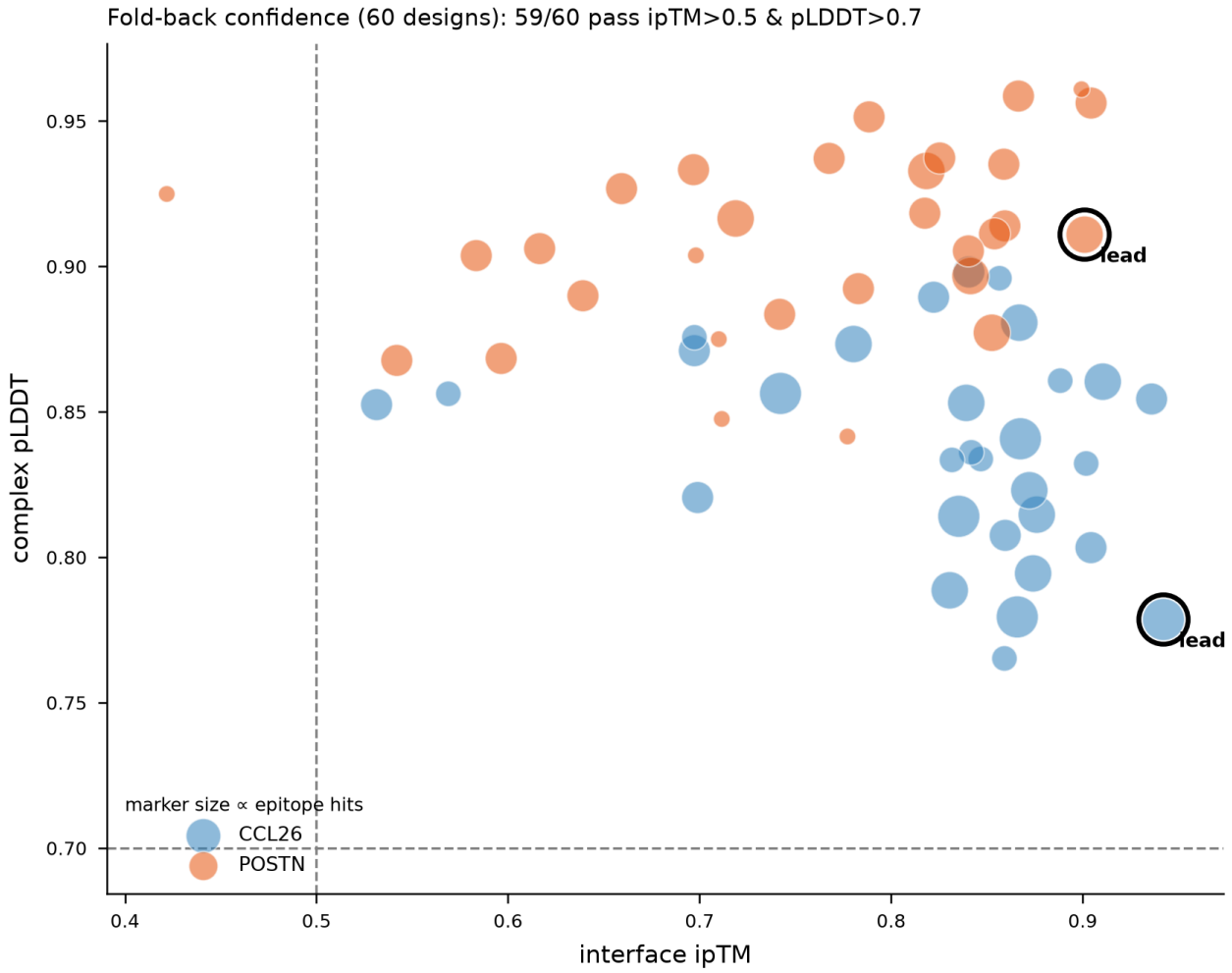

Figure S9. Interface ipTM vs. complex pLDDT for all 60 Boltz-2–folded complexes; marker size proportional to epitope hotspots contacted. 59/60 pass the confidence thresholds (ipTM > 0.5, pLDDT > 0.7; dashed lines). Leads are circled.

Figure S10 — Epitope-coverage distribution

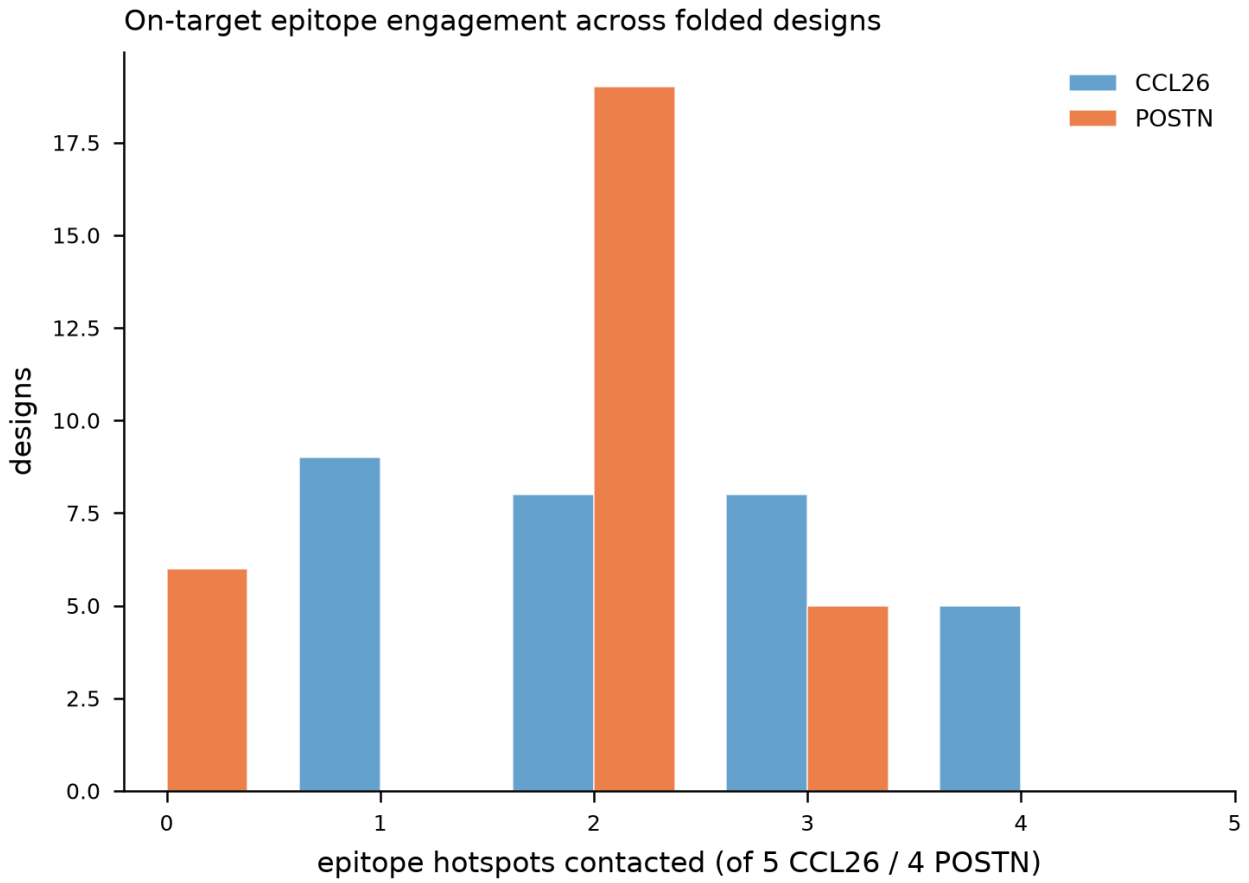

Figure S10. Distribution of the number of epitope hotspots contacted per design (target atom within 5 Å of a binder atom). CCL26 designs engage the epitope more focally (13/30 with ≥3 hits) than POSTN designs (5/30 with ≥3 hits), consistent with the flatter FAS1 integrin face.

Figure S11 — On-target score ranking

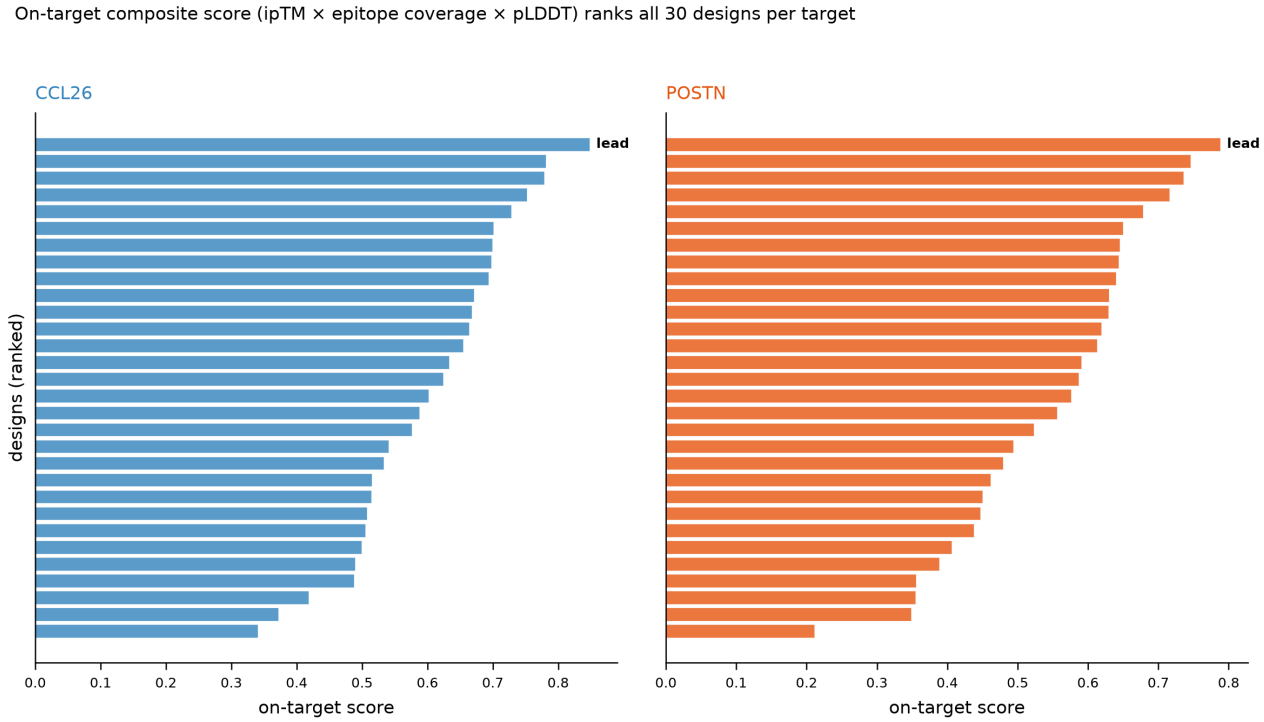

Figure S11. On-target composite score — ipTM × (0.5 + 0.5 × epitope-coverage) × min(pLDDT/0.7, 1) — for all 30 designs per target, ranked. The top-ranked design per target (marked) was selected as the lead.

Figure S12 — Lead interface contact maps

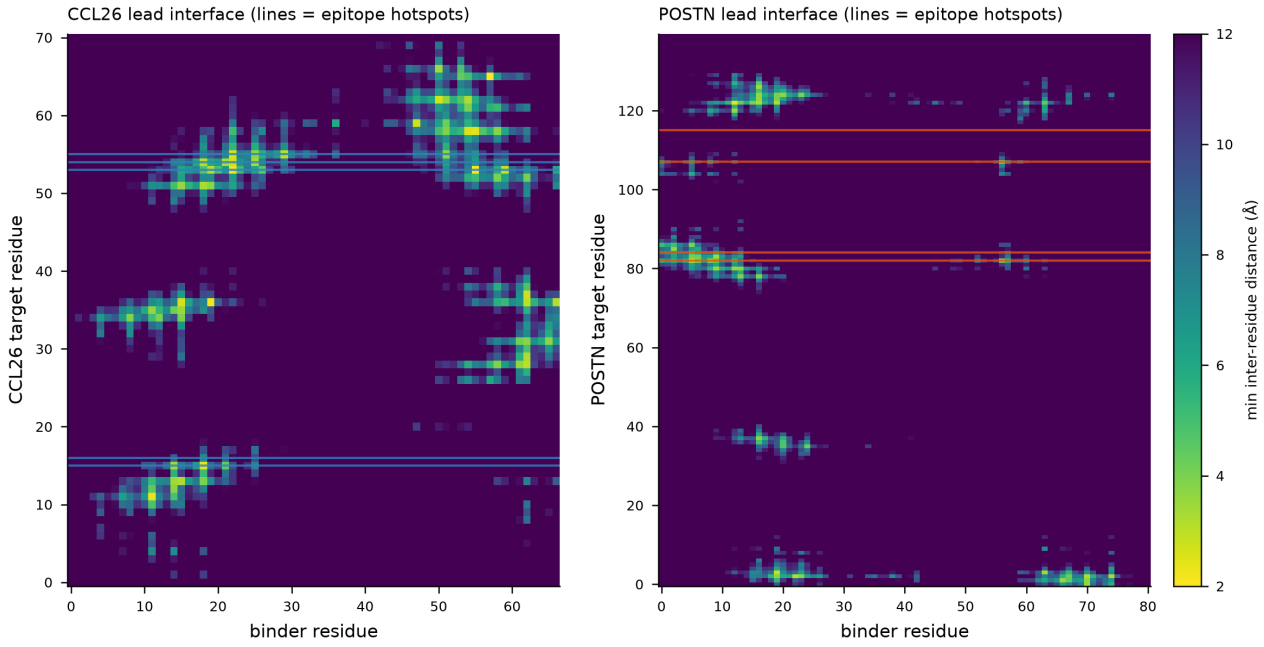

Figure S12. Residue-level minimum inter-chain distance maps for the two lead complexes (target residue × binder residue). Colored horizontal lines mark epitope hotspot rows; the close-contact (yellow) bands align with the hotspots, confirming on-target engagement at the residue level.

Supplementary Tables

- **Table S1 — Meta-analysis (top genes).** Nine-cohort random-effects meta-analysis results (pooled $\log_2$FC, SE, z, p, adjusted p, I^2, k, direction) for the top 60 genes by |z|, including CCL26 and POSTN. `TableS1_meta_top60.csv`
- **Table S2 — Druggability shortlist.** 20-gene omics-derived shortlist with priority score, accessibility, antibody tractability, single-cell cell-types, and suggested modality. `TableS2_druggability_shortlist.csv`
- **Table S3 — Literature × omics cross-reference.** 45 targets with literature abstract/direction counts, trial counts and max phase, and meta-analysis effect sizes. `TableS3_lit_omics_crossref.csv`
- **Table S4 — Fold-back validation (all 60 designs).** Per-design Boltz-2 metrics: ipTM, complex pLDDT, pTM, confidence, interface residue count, epitope hits and coverage, on-target score. `TableS4_foldback_validation.csv`
- **Table S5 — SolubleMPNN summary.** Per-design/temperature sequence summary: sequence count, best/mean MPNN score, mean entropy (240 backbone×temperature groups). `TableS5_mpnn_summary.csv`

Supplementary Media

- **Movie S1 — CCL26 lead complex.** 360° rotation of the CCL26 lead binder–target complex (target grey surface, epitope hotspots blue, binder orange cartoon). `CCL26_binder_complex_360.mp4`
- **Movie S2 — POSTN lead complex.** 360° rotation of the POSTN lead binder–target complex. `POSTN_binder_complex_360.mp4`

Supplementary Methods — key parameters

Stage	Tool	Key parameters
Target prep	Biopython	1G2S chain A (CCL26, 71 aa); 5WT7 chain A (POSTN FAS1-IV, 140 aa); first NMR model, standard residues
Hotspots	Cα spatial compactness	CCL26 [16,17,54,55,56]; POSTN [83,85,108,116]
Backbones	RFdiffusion	Complex/PPI checkpoint; binder 65–90 aa; noise scale 0; 40/target
Sequences	SolubleMPNN (v_48_020)	design chain B, target fixed; 8 seqs × 3 T (0.1/0.2/0.3); 1,920 total
Complexity filter	—	entropy ≥ 2.6; max-run ≤ 5; top-aa ≤ 0.40 (1,074 pass); 30/target selected
Validation	Boltz-2	`--use_msa_server`, 3 recycling steps, 3 diffusion samples; pass = ipTM > 0.5 & pLDDT > 0.7
On-target score	—	ipTM × (0.5 + 0.5 × epitope-coverage) × min(pLDDT/0.7, 1); epitope-coverage = hotspots with target atom < 5 Å of binder
Lead structures	Boltz-2 / Biopython	buried SASA = (SASA_target + SASA_binder − SASA_complex)/2
Rendering	PyMOL + ffmpeg	180-frame 360° turn, libx264 30 fps