A189 · Beyond Proximity: Does AlphaGenome Add Signal Over Splice-Site Annotation in ALS?

Presenter: Arghamitra Talukder — Columbia Univeristy

Authors: Arghamitra Talukder, Alan Kaplan

Session: Day 2 · Fri Oct 2, 2026 · 2:15–4:15 PM

Genomic foundation models are increasingly positioned as general-purpose systems that could reduce reliance on wet-lab validation and accelerate mechanistic discovery. This positioning assumes such models learn principles of regulation rather than memorizing training data. We test that assumption by asking whether AlphaGenome, a state-of-the-art sequence-to-function model, supports variant interpretation in amyotrophic lateral sclerosis (ALS) without any disease-specific training or fine-tuning. We evaluate the model on two complementary tasks. First, we build a matched benchmark from NYGC ALS Consortium splicing QTLs in cervical spinal cord, pairing each lead sQTL with matched control variants and testing discrimination within junction. Second, at the UNC13A locus, we ask whether predictions recover the well-characterized ALS risk mechanism in which intronic variants exacerbate TDP-43-dependent cryptic exon inclusion. AlphaGenome achieves high precision at stringent effect thresholds but recovers only a limited fraction of experimentally supported sQTLs, and predicted effect magnitude shows weak correspondence with measured effect size. Much of the apparent discrimination tracks proximity to annotated splice sites. At UNC13A, the model captures part of the local splice-regulatory signal but does not represent the TDP-43-dependent mechanism itself, and predictions vary substantially with sequence-context window size. These results position current genomic foundation models as high-confidence rule-in tools rather than sensitive screens. For discovery the constraint is sharper: much of the model's discrimination tracks splice-site proximity, so it mostly surfaces candidates that annotation alone would already reach.

Rule-In, Not Discovery: Benchmarking AlphaGenome Against ALS Splicing QTLs

Arghamitra Talukder[1,2], Alan Kaplan[2]
[1]*Computer Science, Columbia University;* [2]*Lawrence Livermore National Laboratory*

Biological foundation models are increasingly presented as general-purpose systems that could reduce wet-lab experimentation and accelerate discovery [1]. This rests on an assumption: that they learn mechanisms of regulation rather than interpolating over training data. Standard benchmarks, evaluated on held-out genomic regions, mostly test whether a model recovers known effects. The claims that matter for discovery are harder—whether it prioritizes candidates annotation would not already nominate, and whether it *explains* why a variant is disease-relevant. We test these by applying AlphaGenome [2] to amyotrophic lateral sclerosis (ALS) without disease-specific training, benchmarking against splicing QTLs measured in ALS cervical spinal cord [3] with junction-matched controls, and against the known TDP-43-dependent cryptic exon mechanism at *UNC13A* [4,5].

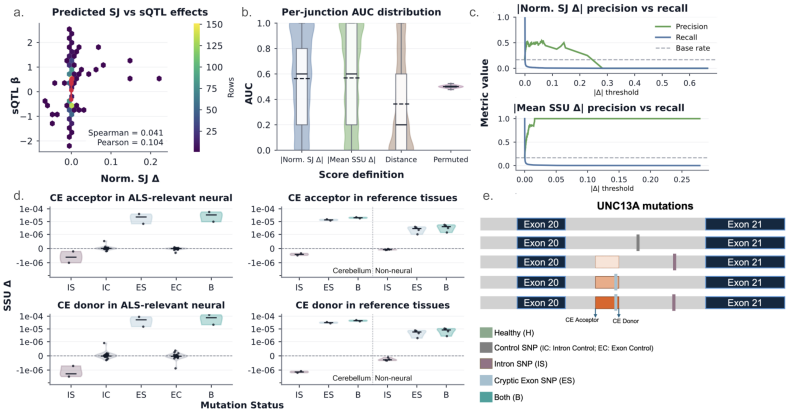

Figure 1: AlphaGenome splicing predictions benchmarked against ALS cervical spinal-cord sQTLs. Predicted splice junction effects show no correspondence with measured sQTL effect sizes (a); per-junction discrimination exceeds both a permutation null and a distance-only baseline but remains weak (b); high precision is reachable only at near-zero recall (c). At *UNC13A*, the cryptic-exon SNP dominates the predicted signal while the intronic GWAS lead receives a near-zero effect of opposite sign (d,e).

Results. AlphaGenome carries information beyond splice-site annotation, but that information is weak, uncalibrated, and tissue-nonspecific. Under our within-junction matched design—sQTL positives at $q < 0.01$ paired 1:5 with LD-pruned controls drawn from the same junction, a 0.167 base rate—a distance-only baseline falls *below* chance, confirming that the design strips out trivial positional cues. Against this, both scores we test, the normalized splice junction change $|\text{Norm. SJ } \Delta|$ and the mean splice-site-usage change $|\text{Mean SSU } \Delta|$, reach a median per-junction AUC of ≈ 0.59 against a permutation null at 0.50 (Fig. 1a,b). The signal is genuine but only marginal, and it is uncalibrated: predicted magnitude does not track measured effect size (Spearman $\rho = 0.041$, Pearson $r = 0.104$), so a larger predicted Δ implies neither a larger nor an equally-signed regulatory effect. Its precision–recall structure is that of a rule-in filter rather than a screen—precision rises above the 0.167 base rate only in a narrow high-magnitude tail and collapses as soon as recall becomes non-trivial (Fig. 1c). At *UNC13A* the model separates both risk variants from their matched controls, yet essentially all predicted signal comes from the SNP inside the cryptic exon; the intronic GWAS lead receives an effect two orders of magnitude smaller and of opposite sign to the measured increase in cryptic-exon inclusion (Fig. 1d,e). The same shifts appear in cerebellum and non-neural tissue as in ALS-relevant neural context, so the prediction reflects a static sequence property rather than the disease-relevant regulatory state. In summary, AlphaGenome identifies a subset of variants with high confidence but fails to recover the majority of true effects, yielding informative predictions at a precision that precludes its use as a sensitive screen.

References. [1] Bommasani *et al.* On the opportunities and risks of foundation models. 2021. [2] Avsec *et al.* AlphaGenome. 2026. [3] NYGC ALS Consortium. Cervical spinal cord sQTL resource. [4] Ma *et al.* TDP-43 represses cryptic exon inclusion in *UNC13A*. *Nature*, 2022. [5] Brown *et al.* TDP-43 loss and ALS-risk SNPs drive mis-splicing of *UNC13A*. *Nature*, 2022.