Verge Labs

TECHNICAL REPORT · JUNE 16, 2026

Introducing vBx-1.0

Verge's multi-modal foundation model for precision neurology

By Verge Labs

FOUNDATION MODELS NEUROLOGY

90% of drugs fail, often because the wrong drug is tested in the wrong patients. Complex diseases rarely boil down to a single clinical diagnosis. What we call Alzheimer's Disease is really many diseases, each driven by a different mix of genetic, cellular, and tissue-level processes. Pool those patients into one trial and a real effect in a subset washes out into noise.

AI has accelerated the early steps of drug discovery. AlphaFold and ESM3 predict molecular structure; Boltz and Chai design small molecules and antibodies; virtual cell models predict how a cell behaves. Every one of them works upstream, at the molecule or the cell. None of them speaks to the patient. Yet most drugs fail in the clinic not because the molecule was wrong, but because we could not predict how the disease would behave in the patients enrolled.

Oncology made this leap towards precision medicine a decade ago: a tumor can be biopsied, profiled, and matched to the right drug. In neuroscience, the disease happens inside the brain, which cannot be sampled in a living patient. So the field has had to rely on proxies: blood, CSF, imaging, EEG. Each captures the consequence of brain disease, not the disease itself. Brain tissue is the molecular ground truth for disease and access to it is the key to unlocking precision neurology.

vBx-1.0: A world model for simulating patient brain biology

vBx-1.0, our Virtual Biopsy model, closes that gap. It reconstructs the molecular state of a patient's brain from measurements you can take from a living person. By fusing peripheral data with learned brain biology, it takes on the hardest questions in translational medicine: how a disease will progress, who will respond to a therapy, and who is at risk of side effects.

vBx-1.0 is a multimodal transformer trained on one of the largest paired brain datasets ever assembled: 3,272 human brains from 1,586 patients across five diseases. The dataset spans brain and blood expression, genetics, and clinical trajectory, with every patient contributing multiple modalities. This initial training set is a subset of the overall 12,000 brain samples across 6,500 patients in Verge's data atlas. As we scale to incorporate further modalities and iterate on our model, we will increase the scale of patients included in the training set.

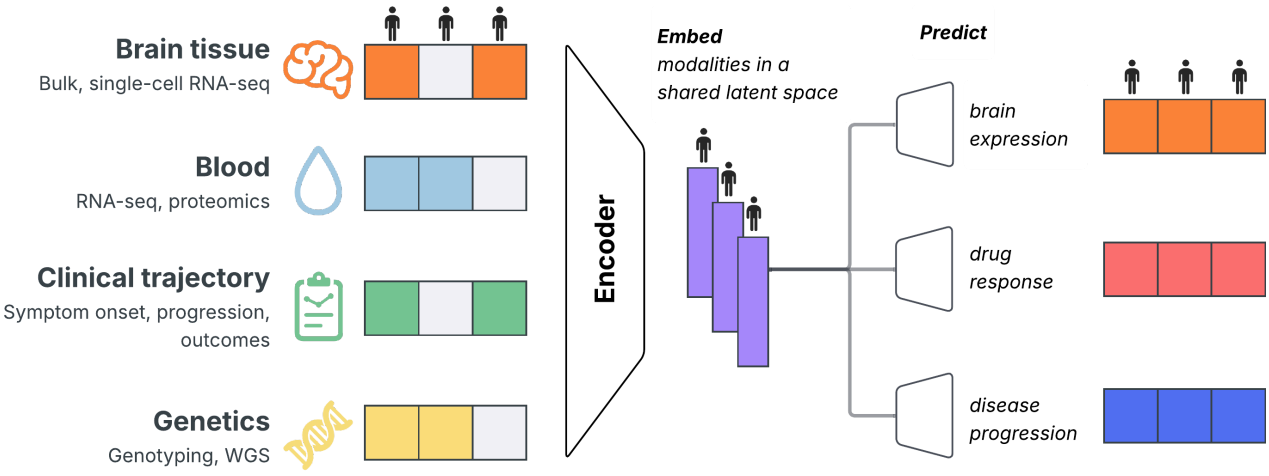

Figure 1. Schematic of Verge's multi-modal foundation model. The model uses a contrastive learning approach that finds synergies between the modalities available for a given patient and generates a unified patient embedding. Patient embeddings are inputs into downstream translational tasks.

The model fuses any combination of a patient's blood RNA-seq, brain RNA-seq, and genetics into a single 512-dimensional embedding that captures their biological state. It is small by design, roughly 90M parameters, and runs on a single GPU. So far the largest gains have come from scaling high-quality paired data rather than compute, though scaling compute is on the roadmap.

vBx-1.0 is trained with a contrastive objective, but not the usual one. Models like CLIP (Contrastive Language-Image Pre-training) keep only what two modalities agree on. That works for image search, but in biology, it throws away the most valuable signal, because each measurement sees something others cannot: genetics speaks to lifetime risk, blood to peripheral immune state, brain omics to what the disease is doing in the tissue itself. So we trained the model to keep both: what each modality knows on its own, and the signal that emerges only when they are combined. We do this by masking out modalities for a patient and forcing the masked view to land close to the full-modality view in latent space.

The payoff is that we can reconstruct the full multimodal picture of a patient from only the modalities that are easy to collect from a living person, namely, blood, genetics, and clinical measures.

That has an immediate practical value. A team with only genotype or blood transcriptomics from a trial can predict brain expression, flag likely responders, and surface patients at risk of adverse events, or fine-tune the embeddings into custom predictors for their specific program. Because the model is small, it embeds 1,000 patients in 16 seconds on a laptop, nimble and accessible enough for any team to use on day one.

We designed our benchmarks around real translational problems, and vBx beats the standard clinical and machine-learning baselines on every one. The headline results are below; each section that follows digs into the data behind them.

The first is reconstructing brain state from blood alone, where vBx is state of the art. It outperforms scVI, a strong deep learning model that beats much larger virtual cell models like scGPT, and a Principal Component Analysis (PCA) baseline standing in for BrainGENIE. In all cases, the models use blood expression as the only input.

Across the case studies that follow, vBx improves on the gold-standard clinical practices in use today for stratifying patients and predicting on-target toxicity. In Parkinson's, enriching a trial for likely responders **cuts required enrollment by 43%** at the same statistical power. In our own ALS trial, screening on predicted baseline biology would have **reduced the dropout rate by 34%.**

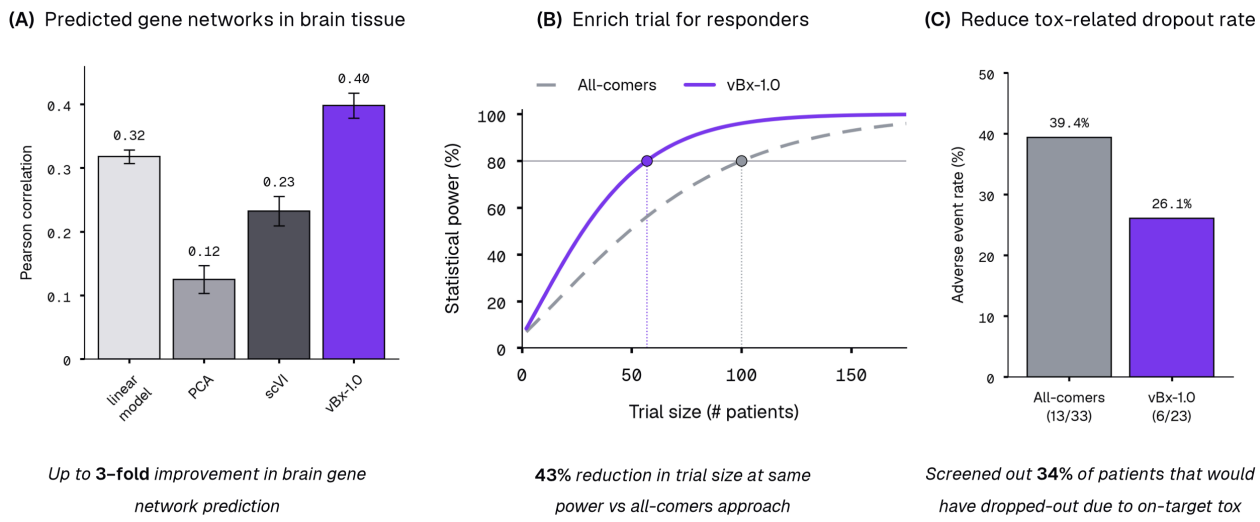

Figure 2. Summary of vBx-1.0's performance on downstream tasks. vBx-1.0 outperforms every SOTA baseline model on gene network prediction in brain tissue (N=26-57 patients) (A). With clinical and omics measurements available at time of enrollment, vBx-1.0 predicts a responder-enriched enrolled population that allows for a 43% reduction in trial size at the same power vs an all-comers approach using L-DOPA responder prediction as a benchmark (B), and screens out 34% of patients that would have dropped out due to an on-target tox adverse event using Verge's VRG50635 trial as a benchmark (C).

Task #1: Predicting brain state from blood

vBx-1.0 sets the state of the art at reconstructing disease-relevant brain biology from blood

Our translational goal ultimately is reading out the disease-relevant pathways active in a patient, since neurodegeneration is driven not by single genes, but by coordinated programs of dysregulation. At Verge, we define these pathways from primary tissue: we profile post-mortem brains across disease and control donors, use our CONVERGE platform to build co-expression networks, and retain the ones that are both strongly dysregulated in disease and enriched for genetic drivers of disease. We call these “disease modules.”

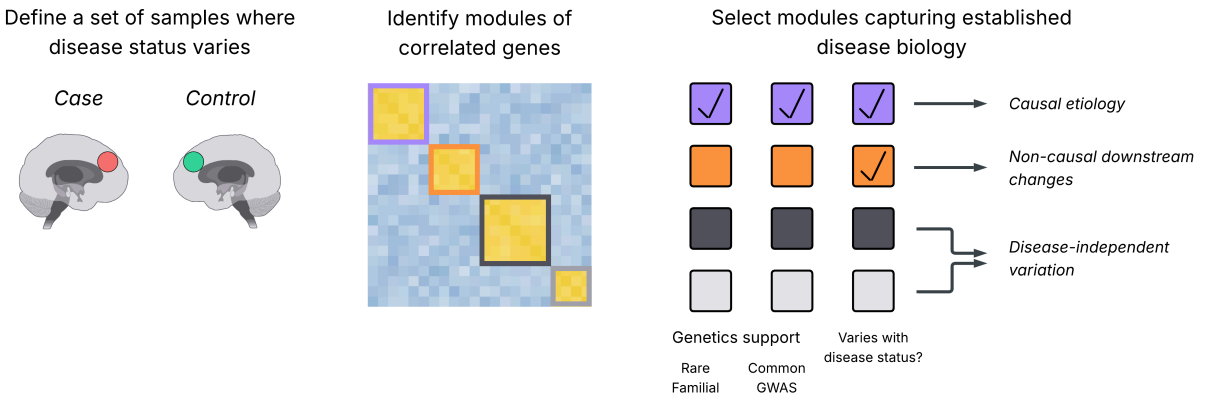

Figure 3. Schematic of how disease pathways are discovered using Verge’s computational biology engine, CONVERGE.

Modules are defined on post-mortem tissue. To use them in living patients, vBx predicts the activity of these same brain-derived modules from blood alone, turning a post-mortem definition into a live, interpretable readout of a patient’s disease biology.

To test this, we held donors with matched blood and brain tissue completely out of training. For each, we fed the model only their blood RNA-seq profiles, predicted the activity of disease-relevant brain modules, and scored the predictions against the donor’s true brain values.

We benchmarked vBx against three approaches that represent how one might attempt this today: a regularized linear model trained directly on blood-brain pairs, capturing whatever is straightforwardly shared between the tissues; PCA, standing in for projection methods such as BrainGENIE; and scVI, a strong deep learning expression model. All three receive the same blood input, so any difference reflects the representation each has learned.

Across diseases and brain regions, vBx outperforms all models. In the substantia nigra, the seat of dopaminergic degeneration in Parkinson's, vBx recovers Parkinson's disease modules with an average Pearson correlation of 0.40 while the baselines top out at 0.27 and scVI collapses to 0.08. For Alzheimer's, it leads in cortex (0.41) and dorsolateral prefrontal cortex (0.49), the latter nearly 60% above the next-best method.

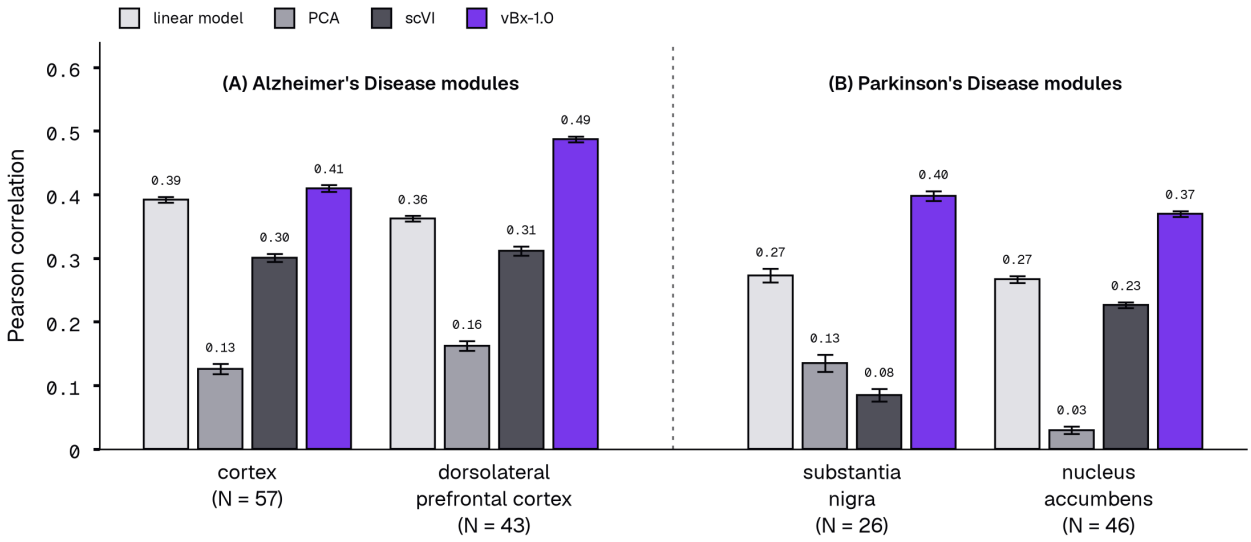

Figure 4. Performance comparison of module activity prediction in various brain regions as predicted from patient embeddings constructed from blood only. Two indications are selected for display from the table displayed above, Alzheimer's modules (A, module count = 102), and Parkinson's modules (B, module count = 113). The performance of vBx-1.0 is displayed alongside three other approaches.

Beating all previous state-of-the-art comparables suggests that by pretraining across modalities, vBx captures signals beyond what conventional approaches can capture from blood alone. This margin matters; blood and brain are weakly coupled in most CNS conditions, so reading brain state from blood is genuinely difficult. That vBx adds a large margin on top means pretraining taught it **something non-obvious about how biology propagates** from one tissue to the other.

And vBx was never trained to map the blood to the brain directly. It learned to do so on its own, by filling in masked modalities during pretraining, the way a language model learns from predicting hidden words. The capability emerged from the objective, not from task-specific supervision.

Together these benchmarks establish the core capability: from blood alone, vBx reconstructs disease-module activity at state-of-the-art accuracy. That is what lets us ask which patients share the biology a drug targets, who will respond, and who is at risk, the questions that decide whether a trial succeeds.

vBx-1.0 accurately reconstructs single brain markers within individual patients

We then zoomed in from modules to single genes. For a held-out patient, we gave the model only blood and asked it to predict the expression of individual brain markers in specific regions, checking each prediction against the patient's matched brain tissue.

We picked four markers spanning different cell types. TH and DAT (SLC6A3) are defining markers of the dopaminergic neurons lost in Parkinson's; GFAP marks astrocytic injury and is one of the most widely used biomarkers in neurology; HSPA8 is a housekeeping gene expressed broadly, including in blood. vBx led the field (mean Pearson 0.25) against a linear model on RNA expression (0.20) and both PCA and scVI (0.14).

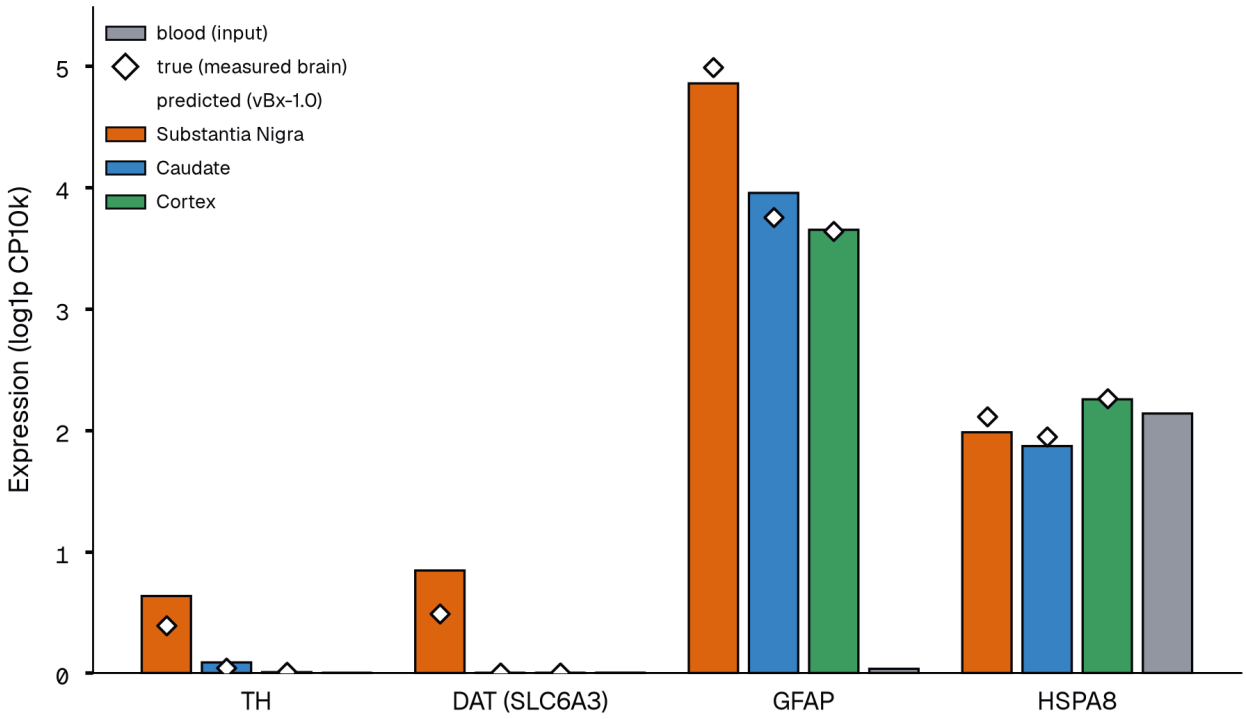

Figure 5. vBx-1.0 predicted (solid bars) and ground truth (diamonds) specific gene expression in different brain regions for an individual donor alongside the donor's input blood transcriptomic profile for the given gene (grey bars).

vBx reconstructs these markers across regions and **even recovers their spatial pattern** (Figure 5): it concentrates TH and DAT in the substantia nigra, where dopaminergic neurons live, while spreading GFAP across all regions, matching how astrocytes are distributed.

TH and DAT are nearly undetectable in blood at either the RNA or protein level, which makes the prediction more striking.

Task #2: Stratifying patients by treatment response

Patient heterogeneity is one of the main reasons drugs fail. A single diagnosis hides many biological subtypes, and a drug that works in one can disappear when its effect is diluted across the others. Safety works the same way: in our own ALS trial, a third of patients were far more sensitive to the drug than healthy volunteers, tracking with disease severity. A trial lives or dies on enrolling the right patients, and the hard part has always been identifying them before treatment.

vBx-1.0 predicts response in Parkinson's patients to standard of care

Levodopa (L-DOPA) is the most effective treatment for Parkinson's motor symptoms, but its benefit varies enormously between patients. Some patients recover most of their motor function on medication; others gain little. The clinician has almost nothing to predict this in advance: response is associated loosely with presentation, disease duration, and age of onset, and genetics contributes signal in specific subgroups such as PRKN, GBA1, and LRRK2, but no biomarker or pharmacogenomic test for levodopa response has reached routine clinical use.

That makes L-DOPA response a demanding benchmark, and Parkinson's offers a straightforward way to score it. The Levodopa Challenge Test measures a patient's motor exam (MDS-UPDRS Part III) in a defined OFF state after overnight washout, then again at peak dose about an hour after levodopa. The OFF-to-ON drop is a direct readout of acute response. Administered annually across a large, deeply characterized cohort, it gives us something rare: a meaningful clinical endpoint, measured consistently and repeated per patient at scale.

We drew that cohort from the Parkinson's Progression Markers Initiative (PPMI), an international study of more than 1,500 Parkinson's patients pairing baseline clinical history, biofluids, and blood transcriptomics with repeated annual challenge tests. We averaged each patient's OFF-to-ON improvement across visits to get a stable picture of how they typically respond (Figure 6). The spread is enormous: the best responders recover 70 to 80% of their motor function, while the weakest barely move.

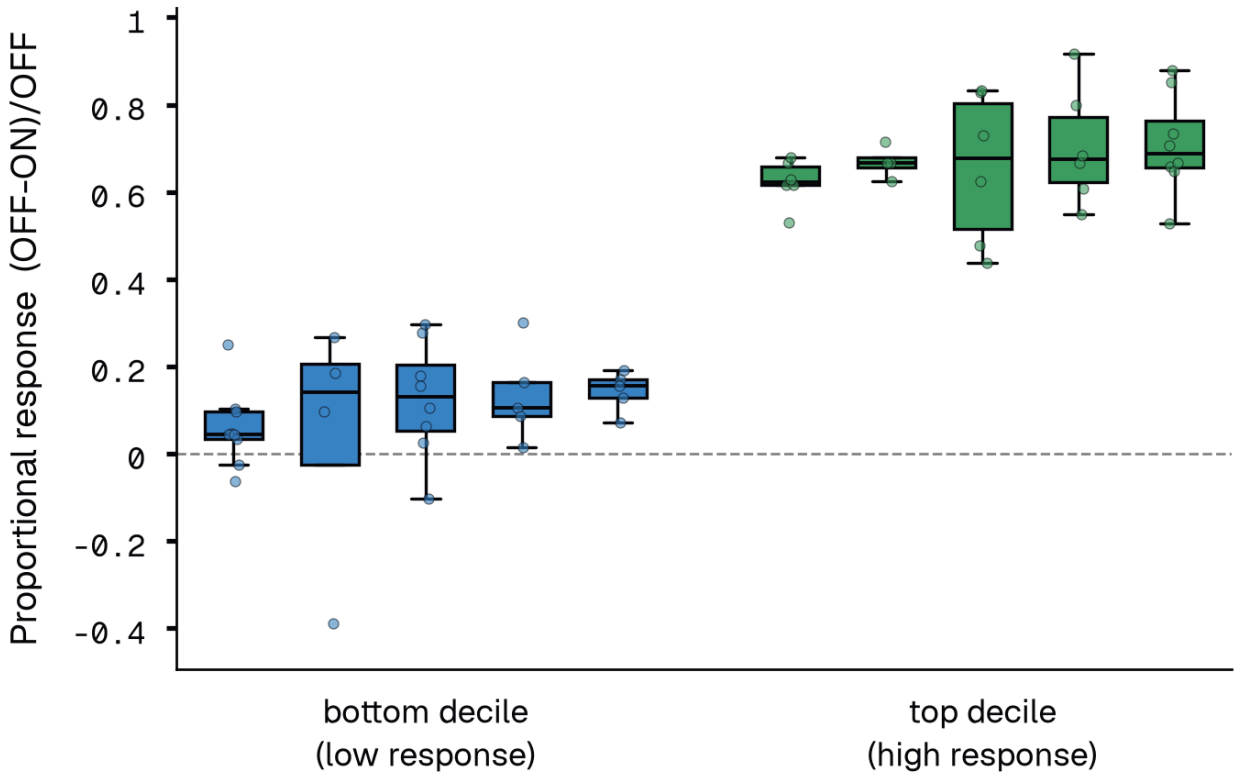

Figure 6. Ten illustrative patients from the PPMI cohort (sampled from the top and bottom decile) and their modeled proportional response to L-DOPA treatment. Each box plot depicts a single patient across their visit history. We estimated each patient's L-DOPA response as the proportional OFF → ON improvement in MDS-UPDRS Part III, and regularized these per-patient estimates with Normal–Normal empirical Bayes, shrinking each patient toward the cohort mean in proportion to the number of paired OFF/ON visits available.

We gave the model a challenging task: using only the information collected at a patient's very first visit, predict how well they will respond to L-DOPA, a treatment typically started years later. Instead of simply sorting patients into "good" or "poor" responders, we asked the model to predict how much each patient would improve.

We trained a small predictor on top of vBx's embeddings, showing it each patient's baseline blood and clinical features paired with their actual L-DOPA response. At test time it saw only a new patient's baseline data and predicted how they would respond.

We compared vBx against four clinical correlates the field uses for treatment decisions: neurofilament-light chain (NfL), UPDRS-III, age at onset, and disease duration (Figure 7).

The results were exciting – vBx-1.0 outperforms every machine-learning approach we tried. Specifically vBx-1.0 beats using clinical features only (0.29 +/- 0.02), as well as replacing vBx-1.0 embeddings with the input blood RNA-seq (0.27 +/- 0.05), PCA (0.30 +/- 0.02), and

scVI (0.30 +/- 0.05). From a translational perspective, vBx-1.0 also beats any single clinical feature a clinician would use today.

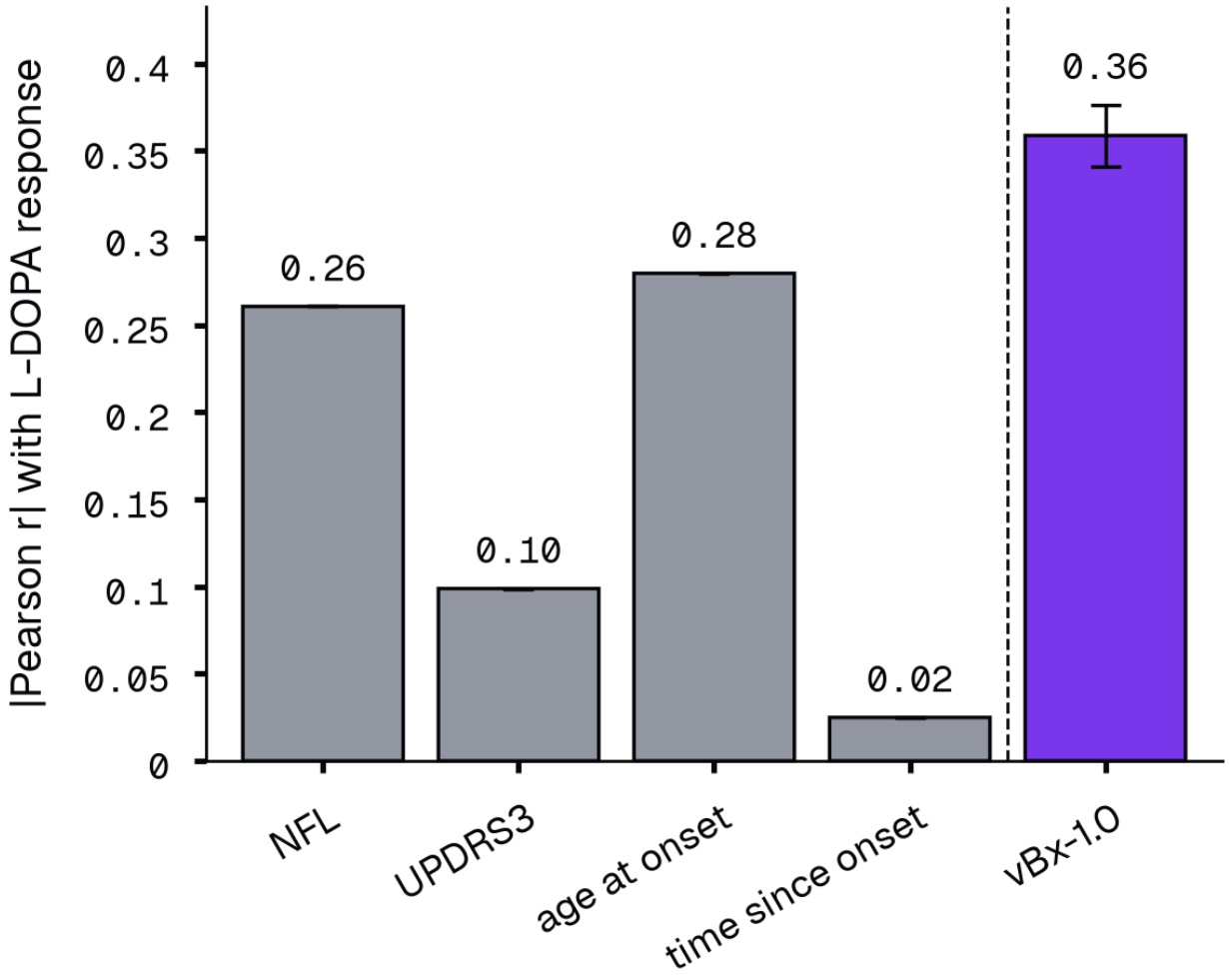

Figure 7. Comparison of various methods for predicting L-DOPA treatment response in Parkinson's disease patients from baseline clinical measurements alone (grey bars), and vBx-1.0's performance in purple. Patients are from the PPMI cohort (N=131). L-DOPA response is defined as described in the main text to be the proportional response.

This task is intrinsically hard: predicting a future drug-induced change in the brain from a blood sample taken before dosing, across the blood-brain barrier, for a phenotype whose usual clinical predictors are nearly uninformative. The payoff is in trial design. Even a modest ability to rank patients by likely response lets a sponsor enrich for responders, concentrating a drug's effect in the right populations and raising statistical power.

Improved response prediction translates to meaningful gains in clinical trial design

A correlation of 0.36 sounds modest. But an enrichment predictor does not need to be right about every patient; it only needs to improve the odds of enrolling responders. The real question is how much that shrinks the trial.

To better simulate a clinical trial with responders and non-responders, we turned the model into a yes/no classifier: from baseline data alone, will this patient meaningfully respond to L-DOPA, meaning at least a 30% motor improvement, the standard clinical bar. vBx identifies responders with 69% accuracy versus 52% for all-comers, a 33% relative gain in responder enrichment.

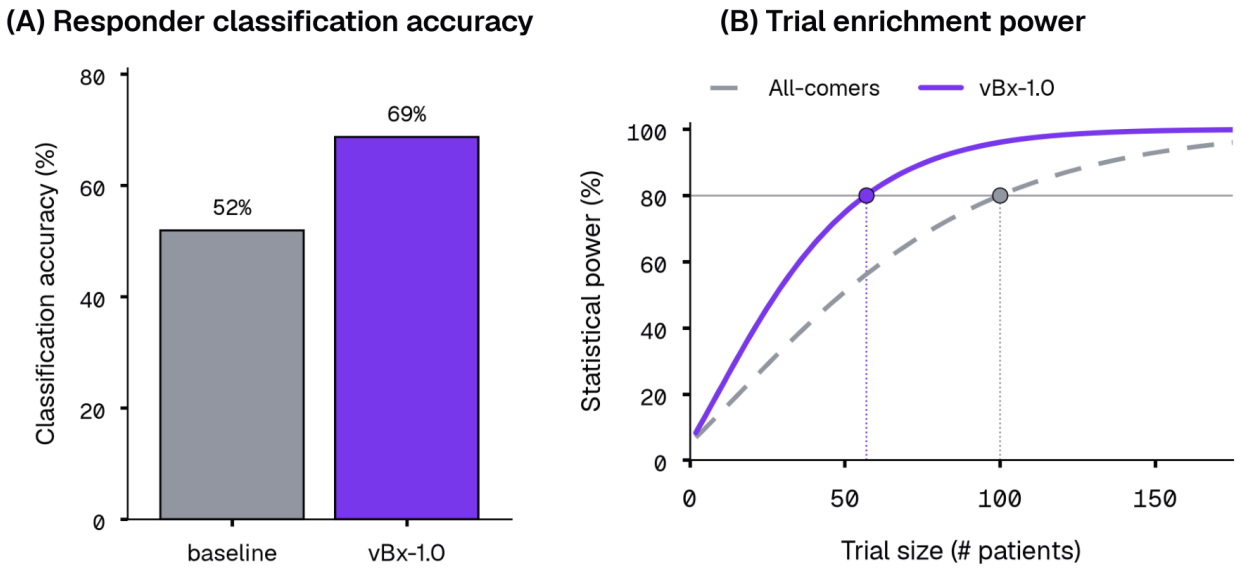

Figure 8. Using vBx-1.0 to enrich a trial for responders yields a 33% relative improvement in the responder fraction (52% to 69%) versus an all-comers enrollment approach, benchmarked using L-DOPA responder prediction in the PPMI cohort. (A) Classification accuracy for identifying responders from baseline data: 52% with all-comers enrollment versus 69% with vBx-1.0. (B) Statistical power as a function of trial size for the enriched versus unenriched cohort, showing that the enriched trial reaches any given power with fewer patients. Assuming non-responders contribute no signal, the average treatment effect is diluted by the responder fraction, giving a standardized effect of $d = p \cdot dR$, where dR is the per-responder effect and p the responder fraction. Required enrollment scales as $n \propto 1/d^2$, so the benefit is the square of the effect-size lift: raising p from 52% to 69% yields $(0.69/0.52)^2 \approx 1.76$-fold greater statistical efficiency, or roughly a 43% reduction in patients enrolled at fixed power.

The surprising part is how much that seemingly modest improvement is worth. Raising the responder share from 52% to 69% doesn't make a trial a little better, it makes it dramatically smaller. The reason is that a trial's ability to detect a drug's effect drops sharply when the group is diluted with people who can't respond, since those patients add noise but no signal.

Clear them out, and the effect becomes much easier to see. Enriching the group this way lets a trial reach the same statistical power with **43% fewer patients** (Figure 8B). In other words for every 100 patients, you would now need only about 57.

vBx-1.0 identifies biological drivers of response heterogeneity

When a model predicts who will respond to therapy, the natural next question is whether you can look inside it and see why. Because the model reconstructs brain biology, we can examine the predicted molecular state of each patient's brain and ask what actually distinguishes responders from non-responders.

Alongside the response prediction itself, we had vBx-1.0 predict the activity of every module and brain region for each PPMI patient. Then we asked whether any of those predicted brain signals tracked L-DOPA response. One module stood out in the predicted substantia nigra. Mapping patients by their vBx profile (Figure 9), the patients who don't respond and the patients with high activity of this module are, strikingly, the same patients.

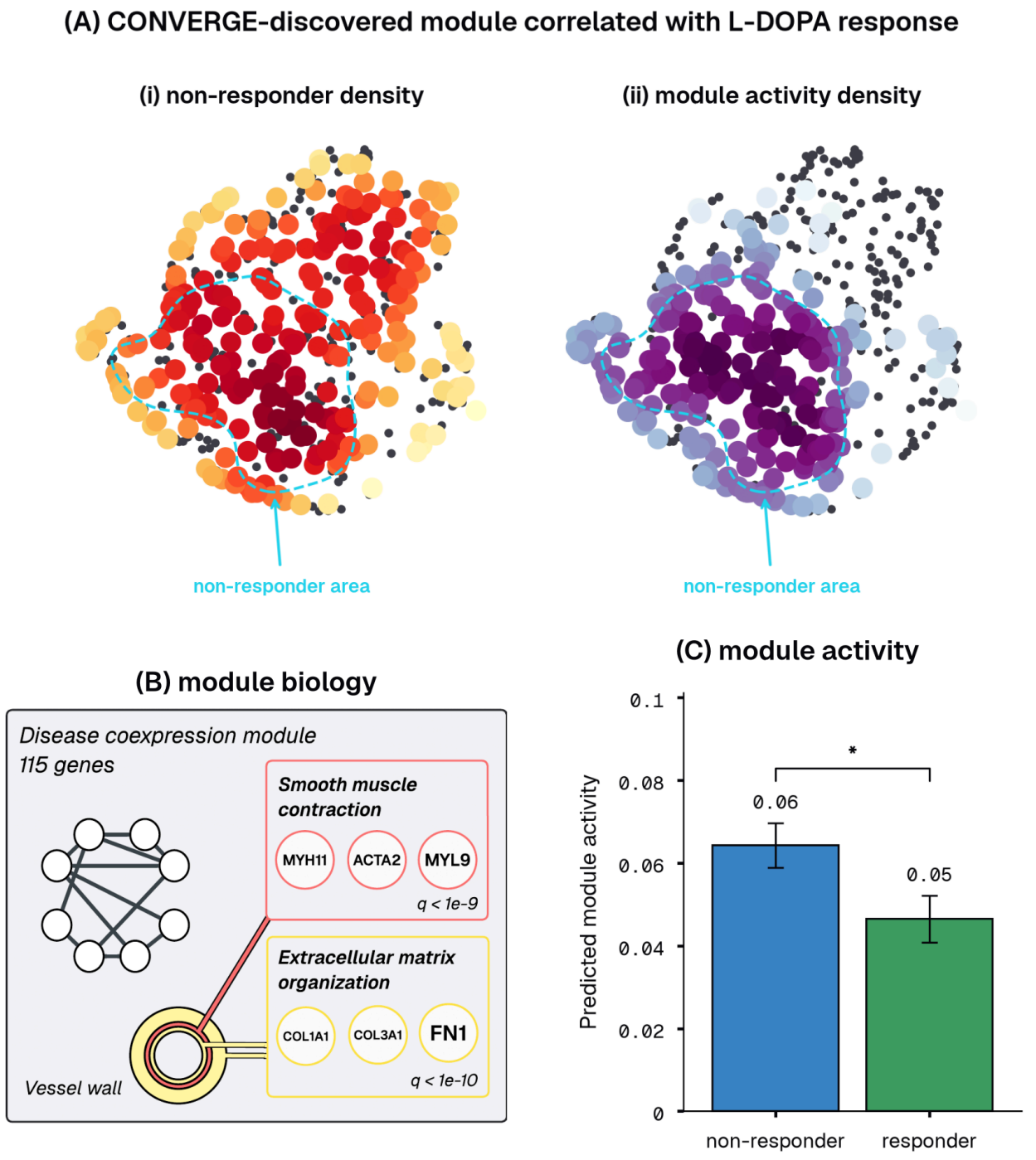

Figure 9. vBx-1.0 patient profiles for the PPMI cohort (n=564), with non-responders defined as patients with less than 30% motor improvement in response to L-DOPA, the standard clinical cutoff. (A) Each dot is one patient, positioned by their vBx profile using UMAP, a projection that places biologically similar patients near each other. Both maps show the same patients in the same positions, colored two different ways. The left map (Ai) is shaded by non-responder density: the warm/red zone inside the dashed outline is where non-responders concentrate. The right map (Aii) shows the same patients colored by predicted activity of the identified module, with dark purple indicating high activity. The dashed line outlines the high-module-activity density core (from ii), overlaid on both panels for reference. The two line up, the region dense with non-responders is the same region where module

activity is high. (B) Summary of the module's biology. (C) The module's predicted substantia nigra activity is significantly higher in non-responders than responders (Mann-Whitney p = 0.033).

When we examined the genes driving this module, we found a vascular signature in which two vessel-wall programs rise together: smooth-muscle contraction and extracellular matrix organization. This pattern is consistent with aging cerebrovasculature, where small-vessel disease and wall remodeling accumulate and reduce cerebral perfusion. In Parkinson's, where the surviving dopaminergic neurons depend on what the vasculature can deliver, this kind of remodeling is positioned to shape not just disease progression but the response to therapy.

The patients who respond least to levodopa are those with the highest predicted vascular burden, where a remodeled wall would both restrict blood flow and lengthen the path the drug must cross to reach surviving terminals. This prediction is not a black box but a specific, biologically coherent vascular program, a mechanistic handle that could become an enrollment biomarker.

Task #3: Anticipating on-target toxicity

The case studies so far rest on strong benchmarks, but none are a real trial. For the hardest test, we turned to one of our own, the trial that motivated this entire model.

In our Phase 1b of VRG50635 (NCT06215755), a brain-penetrant PIKfyve inhibitor for ALS, we ran into an unexpected hurdle. The drug was well-tolerated in healthy volunteers at nearly every dose. But in ALS patients, a third could not tolerate even the lowest dose and discontinued. There were no warning signs in cell lines, animals, or healthy humans – the toxicity was on-target and specific to ALS patient biology. Even within the ALS patient population, different patients responded differently to the same drug. Something in their underlying biology was driving the split, and we didn't know why.

That heterogeneity threatened the trial itself. Losing a third of patients after the first dose shrunk our power to detect an effect in the rest, and we had no way to tell in advance who could tolerate the drug. This is the gap that motivated vBx, and the question here is simple: if we had had vBx at the start of the VRG50635 trial, would it have caught what the trial missed?

vBx-1.0 predicts susceptibility to on-target toxicity in our own clinical trial

This is a genuine **zero-shot** test: vBx saw none of the trial data in pretraining. We pulled baseline whole blood from 33 of the trial's patients, generated RNA-seq, and asked vBx to reconstruct the brain activity of a PIKfyve-anchored module: a phosphoinositide-metabolism program that CONVERGE discovered in ALS brain tissue, built from genes whose proteins interact with PIKfyve, and validated preclinically by showing it drops when PIKfyve is knocked down in patient-derived motor neurons (Figure 10A, Bi). Then we asked: can a patient's baseline brain biology, read from blood, tell us who will tolerate the drug?

The result was clear: patients who went on to have adverse events had lower predicted baseline activity of the PIKfyve module than those who did not (Figure 10Bii).

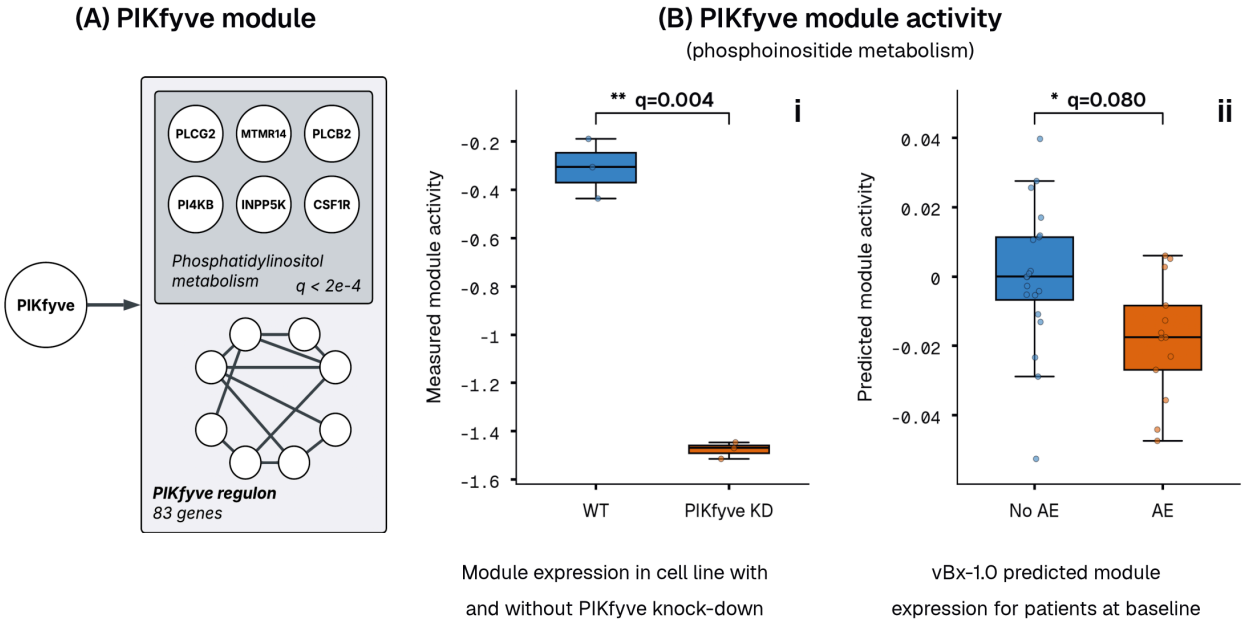

Figure 10. Overview of the CONVERGE-derived PIKfyve module. (A) Network diagram showcasing PIKfyve as a hub gene. (B) Module activity is concordant across the in vitro and patient systems. (Bi) In iPSC-derived motor neurons, the module is down-regulated upon PIKfyve knock-down relative to WT. (Bii) In patients, vBx-1.0-predicted module activity at baseline, before treatment, differentiates those who experienced on-target toxicity, marked by lower activity of this network. AE means "adverse event", and No AE (n=20) are patients who did not experience this vs AE (n=13) did. vBx-1.0-predicted activities from baseline blood, not measured in the brain.

The biology is coherent: PIKfyve inhibition suppresses this module (Figure 10Bi), so patients who already start low have the least reserve to absorb further inhibition, resulting in adverse events (Figure 10Bii). The signal holds after adjusting for NfL and for clinical severity (which we described previously was correlated to adverse events), so it isn't simply re-reading how sick a patient is: the virtual biopsy adds a specific mechanism, read from blood, that explains why those patients had less room to tolerate the drug.

The patients who could not tolerate the drug were, in effect, already running the PIKfyve pathway at its limit. VRG50635 was pushing the same dial, and for the sickest patients there was no room left to push. That is the therapeutic window vBx is built to reveal: where PIKfyve modulation can help, and where baseline reserve is too low to tolerate it.

We wanted to hypothesize about if we had vBx at the time of enrollment and this differentiated module in hand. One simple strategy for using it as an enrollment filter would be to rank all patients by their module activity and then drop the bottom 10 patients; in essence filtering out the patients who would have been most likely to go on to experience on-target tox as the module was most severely down-regulated.

This enrollment screen would have been able to reduce the drop-out rate from 39.4% to 26.1%; a **34% reduction in drop-out rate**.

This closes the gap the PIKfyve trial exposed, tested on the trial that exposed it. The drug was clean in healthy volunteers and stumbled only in patients, which places the liability in disease biology rather than the molecule, and makes it something that should be visible in baseline data. The real confound was not target selection but unmodeled heterogeneity. Stratifying on predicted module activity at enrollment would have recovered the power those dropouts cost us.

Looking ahead

The model gets better with more data

The results so far establish that vBx is working. The more important question for where this goes is why it works, and whether that advantage compounds.

First, brain tissue is what makes the model work (Figure 11A). It is tempting to point AI at whatever data is cheap and abundant. In neuroscience, the most abundant data comes from blood and genetics, which are indirect proxies of disease biology. When we train vBx on “proxy data,” such as blood and genetics, and compare it to the full model that includes the Verge data atlas of multi-modal data from brain tissue, performance on our downstream tasks improves by more than **1.7-fold**. Brain tissue is needed to anchor these datasets to the molecular ground truth of disease.

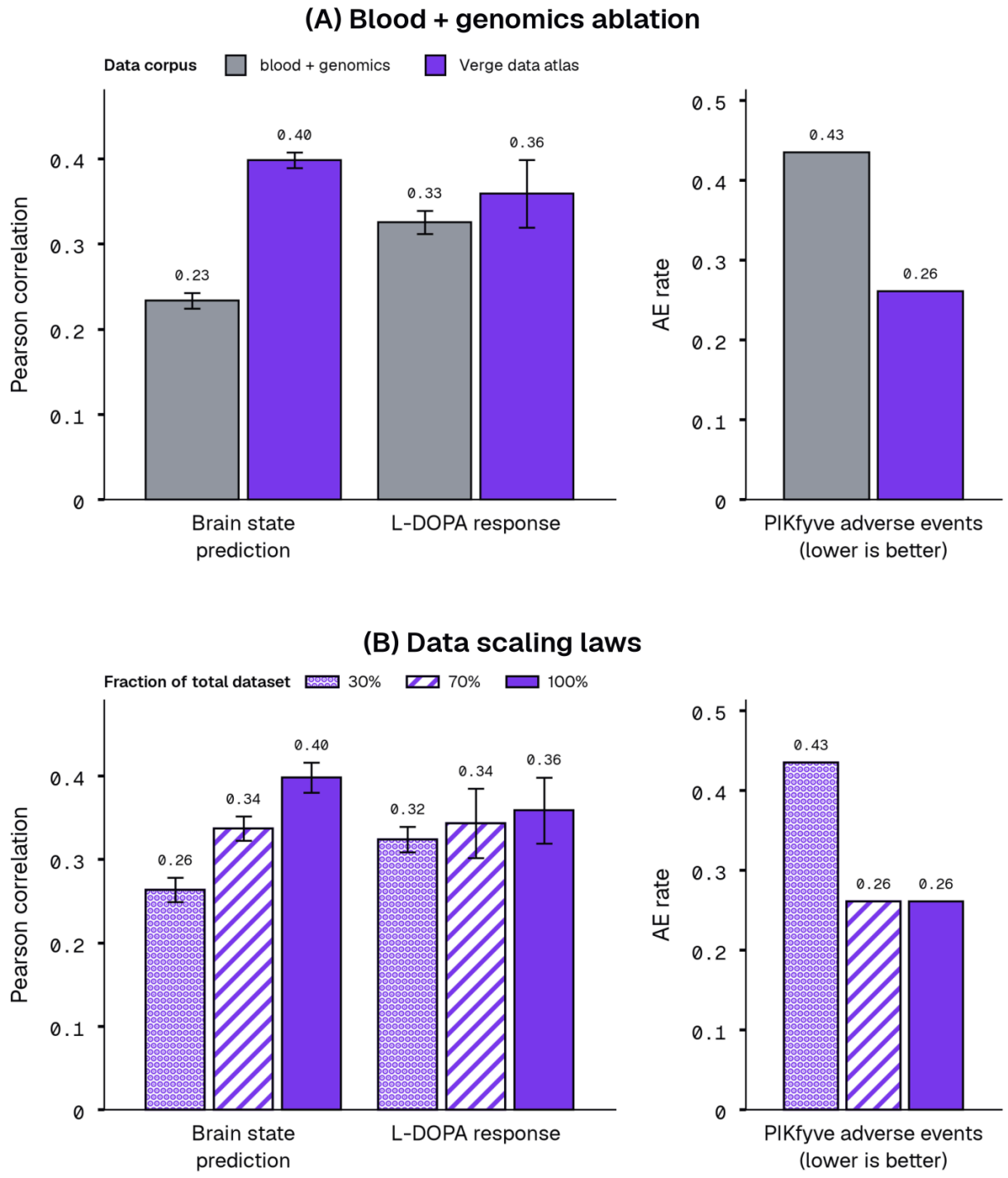

Figure 11. Demonstration of the model's performance on downstream tasks scaling (A) with and without brain data and (B) with additions of more paired patient data points. For the left panels (Pearson correlation), higher is better, and for the right panels (AE rate), lower is better.

Second, vBx obeys scaling laws: more data reliably results in better performance, with enormous headroom still ahead. Holding the validation and test sets fixed, we trained on 30%, 70%, and 100% of our patients, and performance rose at every step across every task

with no sign of saturation (Figure 11B). Unlike language models, the gains come not from parameters but from adding patients and modalities. We are at roughly 90M parameters on a single GPU and have not yet even begun to scale compute.

What's next

Our version 1.0 focuses on genomics, transcriptomics, and clinical data. Proteomics and imaging are next, along with longitudinal modeling that runs a patient forward in time to predict their disease trajectory and how an intervention might change it.

Taken together, CONVERGE and vBx-1.0 do something no platform in neuroscience has done: discover the target and identify the patients whose biology will respond to it, a companion for both discovery and trial design.

We're actively looking for partners who are interested in running vBx-1.0 on their clinical trial data. If you have a clinical trial with heterogeneity you'd like to better understand, or if you are interested in exploring a trial prospectively, please reach out to partnerships@vergelabs.com.