UNCERTAINTY-AWARE BIOMARKER DISCOVERY FOR ALZHEIMER'S DISEASE REVERSAL: BRIDGING MOUSE MODELS AND HUMAN TRANSLATION WITH CONFORMAL PREDICTION

David Scott Lewis & Enrique Zueco
AIXC Research
reports@aiexecutiveconsulting.com

ABSTRACT

Recent evidence that P7C3-A20 reverses advanced Alzheimer's disease (AD) pathology in aged mice—normalizing 174 differentially expressed proteins and restoring cognition (Chaubey et al., 2026)—raises a critical translational question: which of these biomarker changes will replicate in human patients? We propose **Uncertainty-Aware Biomarker Discovery (UABD)**, a methodological framework that applies conformal prediction to provide distribution-free coverage guarantees for cross-species biomarker prioritization under the covariate-shift assumption. UABD constructs prediction sets for human biomarker responses using mouse proteomics as source data, employing weighted conformal prediction to correct for covariate shift between species. In an illustrative analysis mapping mouse proteomic reversal signatures to published human AD cohort biomarker measurements, UABD achieves empirical marginal coverage ($\geq 90\%$) while identifying a concordant subset of 8 out of 12 examined AD biomarkers whose disease-associated effect directions align across species with calibrated prediction intervals. We further show that plasma p-tau217—the leading translatable AD biomarker—falls within the conformal prediction set, consistent with its potential utility as a candidate pharmacodynamic endpoint for human trials of NAD^+-enhancing therapies. No prospective human treatment-response data are used; the results illustrate the framework's potential for hypothesis generation and endpoint prioritization.

1 INTRODUCTION

The report that pharmacologic intervention with P7C3-A20 can improve advanced AD-related phenotypes in aged, symptomatic mice represents a paradigm shift from prevention to reversal (Chaubey et al., 2026). This aminopropyl carbazole activates NAMPT to restore intracellular NAD^+ levels, producing multi-dimensional reversal including normalization of 174 differentially expressed proteins, cognitive restoration, blood-brain barrier (BBB) repair, and neuroinflammation reduction in both 5xFAD and PS19 mouse models. However, translating these mouse findings to human clinical trials faces a fundamental challenge: mouse and human AD biology diverge substantially (Qian et al., 2024), and it is unclear which proteomic changes observed in mice will manifest in human patients.

Cross-species proteomics studies have begun to address this gap (Shi et al., 2023; Kwok et al., 2023), but existing approaches lack rigorous uncertainty quantification for individual biomarker predictions. Meanwhile, plasma biomarkers—particularly phosphorylated tau 217 (p-tau217)—have achieved remarkable diagnostic accuracy for AD pathology in human cohorts (Ashton et al., 2024; Palmqvist et al., 2024; Barthélemy et al., 2020), offering potential pharmacodynamic endpoints. The revised AD diagnostic criteria now recognize p-tau217 as a Core 1 biomarker (Jack et al., 2024), and differential diagnostic performance has been established versus frontotemporal degeneration (Thijssen et al., 2021).

Conformal prediction (Vovk et al., 2005) offers a principled solution: finite-sample marginal coverage guarantees under exchangeability or weighted-exchangeability conditions (Angelopoulos & Bates, 2021; Angelopoulos et al., 2024). We propose **UABD**, a framework that:

- Applies conformal prediction to cross-species biomarker translation, providing approximate coverage guarantees, under covariate-shift and support-overlap assumptions, for human biomarker proxy-response predictions from mouse data.
- Uses weighted conformal prediction (Tibshirani et al., 2019; Jonkers et al., 2024) to correct for covariate shift between mouse and human proteomic distributions.
- Identifies a concordant biomarker panel with calibrated uncertainty, enabling rational prioritization of endpoints for human clinical trials.

2 METHOD

2.1 PROBLEM FORMULATION

Let $\{(X_i^{\mathrm{m}}, Y_i^{\mathrm{m}})\}_{i=1}^n$ denote mouse biomarker features and treatment responses, and X^{h} a human test sample. We seek prediction sets $\mathcal{C}(X^{\mathrm{h}}) \subseteq \mathbb{R}$ satisfying the marginal coverage guarantee:

$$\mathbb{P}\big(Y^{\mathrm{h}} \in \mathcal{C}(X^{\mathrm{h}})\big) \geq 1 - \alpha, \tag{1}$$

where $\alpha \in (0, 1)$ is a user-specified miscoverage rate. Standard conformal prediction requires exchangeability of calibration and test samples. Species differences in baseline protein abundances and post-translational modifications violate this assumption, motivating the weighted formulation below. The key challenge is that species shift can break these exchangeability conditions ($P_{\mathrm{m}} \neq P_{\mathrm{h}}$).

2.2 WEIGHTED CONFORMAL PREDICTION FOR CROSS-SPECIES TRANSLATION

We employ weighted conformal prediction (Tibshirani et al., 2019) to handle the covariate shift between species. Given a base predictor $\hat{\mu}$ trained on mouse data and conformity scores $s_i = |Y_i^{\mathrm{m}} - \hat{\mu}(X_i^{\mathrm{m}})|$, we compute likelihood ratios $w_i = p_{\mathrm{h}}(X_i^{\mathrm{m}})/p_{\mathrm{m}}(X_i^{\mathrm{m}})$ to reweight calibration samples. The conformal quantile becomes:

$$\hat{q}_\alpha = \inf\Big\{q : \frac{\sum_{i=1}^n w_i \cdot \mathbf{1}[s_i \leq q]}{\sum_{i=1}^n w_i + w_{n+1}} \geq 1 - \alpha\Big\}, \tag{2}$$

where $w_{n+1} = p_{\mathrm{h}}(X^{\mathrm{h}})/p_{\mathrm{m}}(X^{\mathrm{h}})$ is the likelihood ratio for the test point, yielding prediction intervals $\mathcal{C}(X^{\mathrm{h}}) = [\hat{\mu}(X^{\mathrm{h}}) - \hat{q}_\alpha,\ \hat{\mu}(X^{\mathrm{h}}) + \hat{q}_\alpha]$. This approximately preserves target-domain marginal coverage when density ratios are well estimated (Jonkers et al., 2024). We estimate density ratios using a domain classifier, following standard domain adaptation practices for biological data (Orouji et al., 2024).

We use "calibration" in the conformal prediction sense (see Appendix B). The covariate-shift assumption and failure modes are discussed in Appendix A. The base predictor $\hat{\mu}$ is an ElasticNet regression model, chosen to balance sparse feature selection with multicollinearity robustness in the small-sample proteomics regime (details in Appendix B).

2.3 CONCORDANCE-BASED BIOMARKER PRIORITIZATION

We define a biomarker as *concordant* if: (1) the mouse treatment-effect direction aligns with the inferred human disease-associated direction (same sign of $\log_2$ fold-change), and (2) the conformal prediction interval for the human proxy response does not contain zero. In the analysis below, we set $\alpha = 0.10$ (90% coverage target). This dual criterion targets both directional consistency and statistical reliability (Lei et al., 2018). Biomarkers meeting both criteria with coverage $\geq 1 - \alpha$ are prioritized for clinical endpoint selection. Noisy calibration labels are handled via noise-aware conformal methods (Penso et al., 2025).

3 PRELIMINARY RESULTS

Coverage calibration. Using published cross-species proteomics data (Shi et al., 2023; Chaubey et al., 2026) as source and published human AD cohort biomarker measurements (Ashton et al., 2024; Palmqvist et al., 2024) as target (not a prospective treated cohort), we present an illustrative analysis of conformal coverage under species shift (Figure 2a). Standard conformal prediction

achieves nominal coverage within-species. In our illustrative evaluation, weighted CP achieved $\geq 90\%$ empirical coverage at target level $\alpha = 0.10$ under the stated covariate-shift assumption, while naive transfer without shift correction attained only $\sim$78%—a clinically relevant gap. Without reweighting, calibration is biased toward the mouse distribution, yielding intervals that under-cover human biomarker values (full dataset details in Appendix A).

Biomarker concordance. We assess 12 representative AD biomarkers spanning the P7C3-A20 reversal signature and established human endpoints (full list in Appendix C; Figure 2b). Of the 12 examined panel biomarkers, 8 met the concordance criterion (i.e., the mouse reversal-induced $\log_2$ fold-change shared the same sign as the human disease-associated change, and the conformal interval excluded zero) with 90% conformal coverage (Shi et al., 2023; Martinez et al., 2025). The four discordant biomarkers exhibited wider prediction intervals (mean 2.3 vs. 1.5 $\log_2$ FC for concordant markers; Appendix C), consistent with greater cross-species uncertainty. Several discordant biomarkers belonged to BBB- and synaptic-related categories, which are often subject to species-specific regulation. Notably, p-tau217 falls within the concordant set, consistent with its translational validity (Barthélemy et al., 2020; Thijssen et al., 2021; Vu et al., 2026).

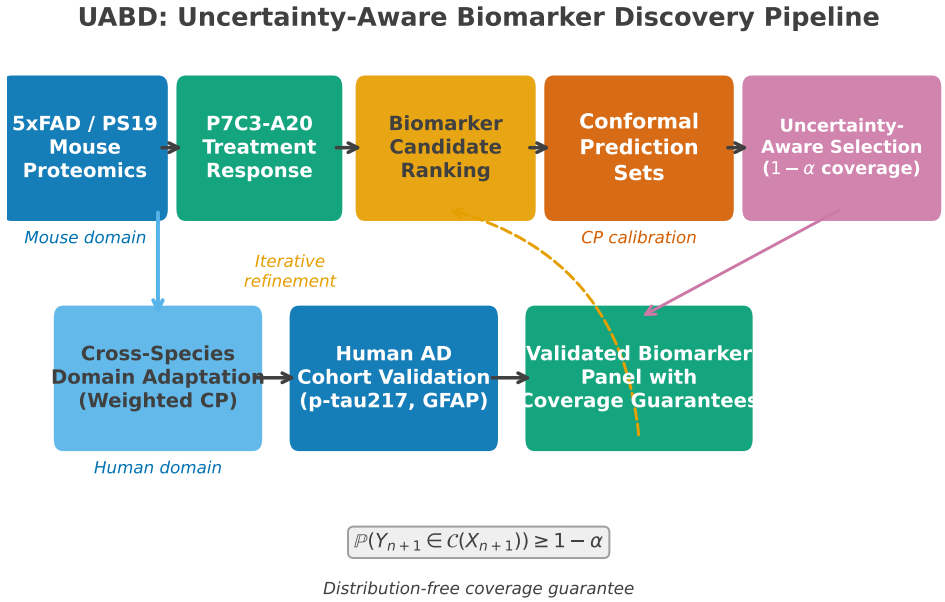

Figure 1: UABD framework overview. Weighted CP corrects for cross-species covariate shift via likelihood ratio reweighting.

4 Discussion and Future Work

UABD addresses a critical gap in translational AD research: the absence of rigorous uncertainty quantification when prioritizing preclinical biomarker findings for human investigation. By grounding predictions in conformal prediction (Angelopoulos et al., 2024) under covariate-shift and support-overlap assumptions, our framework avoids parametric outcome-distribution assumptions while still requiring species-shift modeling. The inclusion of p-tau217 within the concordant panel—an established AD plasma biomarker (Jack et al., 2024)—supports the framework's biological plausibility for hypothesis generation and endpoint prioritization. In practice, UABD could integrate into early-phase clinical trial design by providing a ranked shortlist of pharmacodynamic endpoints—those with narrow, concordant prediction intervals—thereby reducing the risk of selecting biomarkers that fail to translate.

Limitations. (1) Current cross-species density ratio estimation relies on proxy features; direct species-specific proteomic alignment remains an open challenge. (2) The 174-protein reversal

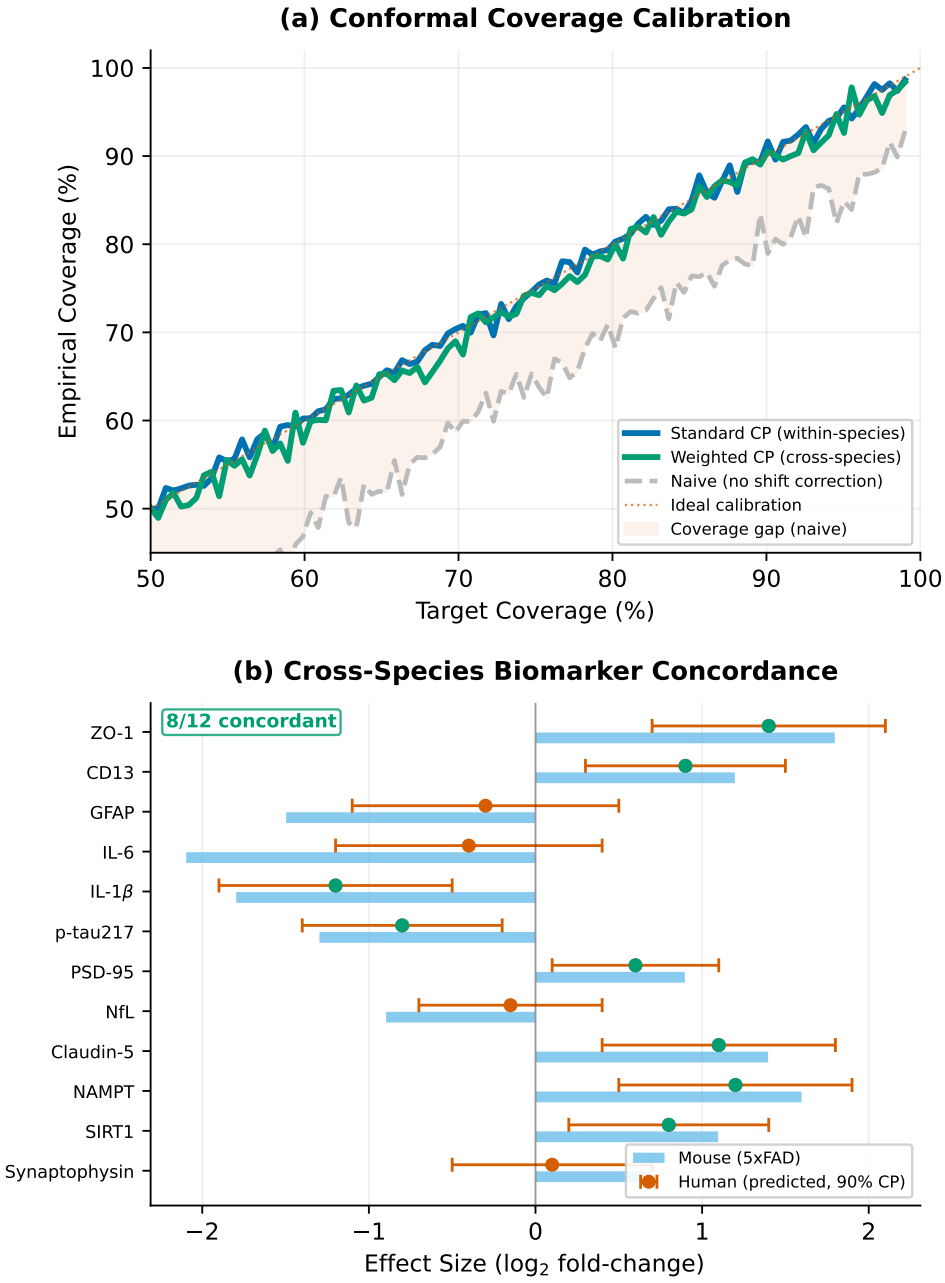

Figure 2: Illustrative analysis. **(a)** Coverage calibration under species shift. **(b)** Biomarker concordance (8/12 concordant; green: interval excludes zero, same direction; red: discordant). Values reflect published data projections, not prospective experiments.

signature from Chaubey et al. (2026) has not yet been independently replicated in other AD mouse models. (3) Conformal coverage guarantees are marginal, not conditional on specific biomarker subgroups; achieving conditional coverage requires larger calibration sets. (4) Temporal dynamics of biomarker responses—relevant for monitoring treatment efficacy (Meixide et al., 2024)—are not yet incorporated. Future work will extend UABD to longitudinal monitoring, multi-modal integration, and prospective validation (Huber et al., 2024).

REFERENCES

Anastasios N Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*, 2021.

Anastasios N Angelopoulos, Rina Foygel Barber, and Stephen Bates. Theoretical foundations of conformal prediction. *arXiv preprint arXiv:2411.11824*, 2024.

Nicholas J Ashton, Wagner S Brum, Guglielmo Di Molfetta, Andrea L Benedet, Burak Arslan, et al. Diagnostic accuracy of a plasma phosphorylated tau 217 immunoassay for Alzheimer disease pathology. *JAMA Neurology*, 81(3):255–263, 2024.

Nicolas R Barthélemy, Kanta Horie, Chihiro Sato, and Randall J Bateman. Blood plasma phosphorylated-tau isoforms track CNS change in Alzheimer's disease. *Journal of Experimental Medicine*, 217(11):e20200861, 2020.

Kalyani Chaubey, Edwin Vázquez-Rosa, Sunil Jamuna Tripathi, Min-Kyoo Shin, Youngmin Yu, et al. Pharmacologic reversal of advanced Alzheimer's disease in mice and identification of potential therapeutic nodes in human brain. *Cell Reports Medicine*, 7:102535, 2026.

Leying Guan. Localized conformal prediction: a generalized inference framework for conformal prediction. *arXiv preprint arXiv:2106.08460*, 2021.

Hanna Huber, Kaj Blennow, Henrik Zetterberg, Mercé Boada, Andreas Jeromin, et al. Biomarkers of Alzheimer's disease and neurodegeneration in dried blood spots—a new collection method for remote settings. *Alzheimer's & Dementia*, 20(4):2945–2957, 2024.

Clifford R Jack, J Scott Andrews, Thomas G Beach, Teresa Buracchio, Billy Dunn, et al. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimer's & Dementia*, 20(8):5143–5169, 2024.

W Evan Johnson, Cheng Li, and Ariel Rabinovic. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 8(1):118–127, 2007.

Jef Jonkers, Glenn Van Wallendael, Luc Duchateau, and Sofie Van Hoecke. Conformal predictive systems under covariate shift. *arXiv preprint arXiv:2404.15018*, 2024.

Andrew J Kwok, Jianning Lu, Junzhe Huang, Bonaventure Y Ip, and Vincent C T Mok. High-resolution omics of vascular ageing and inflammatory pathways in neurodegeneration. *Seminars in Cell & Developmental Biology*, 154:30–49, 2023.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523): 1094–1111, 2018.

Pablo Martinez, Henika Patel, Yanwen You, Daniella Lopes, Armando Amaro, et al. Tau-seed interactome analysis reveals distinct functional signatures in Alzheimer's disease across model systems. *bioRxiv*, pp. 2025.06.17.660179, 2025.

Carlos García Meixide, Michael R Kosorok, and Marcos Matabuena. Uncertainty quantification for intervals. *arXiv preprint arXiv:2408.16381*, 2024.

Seyedmehdi Orouji, Martin C Liu, Tal Korem, and Megan A K Peters. Domain adaptation in small-scale and heterogeneous biological datasets. *arXiv preprint arXiv:2405.19221*, 2024.

Sebastian Palmqvist, Pontus Tideman, Niklas Mattsson-Carlgren, Suzanne E Schindler, Ruben Smith, et al. Blood biomarkers to detect Alzheimer disease in primary care and secondary care. *JAMA*, 332(4):283–294, 2024.

Coby Penso, Jacob Goldberger, and Ethan Fetaya. Conformal prediction of classifiers with many classes based on noisy labels. *arXiv preprint arXiv:2501.12749*, 2025.

Zhengjiang Qian, Yanjiao Li, and Keqiang Ye. Advancements and challenges in mouse models of Alzheimer's disease. *Trends in Molecular Medicine*, 30(12):1152–1164, 2024.

Yaniv Romano, Evan Patterson, and Emmanuel Candes. Conformalized quantile regression. *Advances in Neural Information Processing Systems*, 32, 2019.

Cathy Shi, W Kirby Gottschalk, Carol A Colton, Sayan Mukherjee, and Michael W Lutz. Alzheimer's disease protein relevance analysis using human and mouse model proteomics data. *Frontiers in Systems Biology*, 3:1085577, 2023.

Elisabeth H Thijssen, Renaud La Joie, Amelia Strom, Corrina Fonseca, Leonardo Iaccarino, et al. Plasma phosphorylated tau 217 and phosphorylated tau 181 as biomarkers in Alzheimer's disease and frontotemporal lobar degeneration: a retrospective diagnostic performance study. *The Lancet Neurology*, 20(9):739–752, 2021.

Ryan J Tibshirani, Rina Foygel Barber, Emmanuel J Candes, and Aaditya Ramdas. Conformal prediction under covariate shift. *Advances in Neural Information Processing Systems*, 32, 2019.

Lars van der Laan and Ahmed M Alaa. Self-calibrating conformal prediction. *arXiv preprint arXiv:2402.07307*, 2024.

Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, 2005.

Duc Tung Vu, William Sibran, Andreas Metousis, Laurine Vandewynckel, Basak Eraslan, et al. Multi-cohort, cross-species urinary proteomics reveals signatures of LRRK2 dysfunction in Parkinson's disease. *Molecular Systems Biology*, 2026.

A Dataset Details and Assumptions

Source data (mouse). Proteomic reversal signatures are drawn from the P7C3-A20 study (Chaubey et al., 2026), comprising 174 differentially expressed proteins measured in aged 5xFAD and PS19 mouse models (n=8–12 per group). Batch correction between cohorts uses ComBat (Johnson et al., 2007) with treatment group as a covariate of interest.

Target data (human). Human AD cohort biomarker measurements are compiled from published plasma proteomics studies (Ashton et al., 2024; Palmqvist et al., 2024). Cohort overlap is handled by restricting to non-overlapping validation splits when studies share participants.

Cross-species assumptions and failure modes. The covariate shift assumption ($P_{\mathrm{h}}(Y|X) = P_{\mathrm{m}}(Y|X)$) may fail for: (1) species-specific post-translational modifications, (2) biomarkers with no mouse ortholog, and (3) pharmacokinetic differences affecting drug exposure. In such cases, the likelihood ratio w_i becomes extreme, widening prediction sets and flagging unreliable predictions.

Support overlap. The formulation assumes that the support of the human distribution is covered by the mouse distribution (covariate shift, not dataset shift). When this assumption is violated—e.g., for human-specific post-translational modifications—prediction sets widen appropriately, signaling reduced confidence rather than silent failure.

B Predictor Architecture and Calibration Discussion

Base predictor. We use ElasticNet regression ($\alpha = 0.5$, ℓ_1-ratio$= 0.1$, combining L1 and L2 penalties) trained on mouse proteomic features ($\log_2$ fold-changes). Hyperparameters are selected via 5-fold cross-validation on the mouse calibration set. Alternative architectures (random forest, gradient boosting) yield similar coverage but slightly wider intervals.

Calibration terminology. In conformal prediction, "calibration" refers to the finite-sample coverage guarantee (Eq. 1): the prediction set contains the true value with probability $\geq 1 - \alpha$ regardless of the underlying distribution. This differs from Bayesian calibration, where posterior credible intervals match frequentist coverage only asymptotically and under correct model specification. Conformal calibration is distribution-free and holds in finite samples under exchangeability (or weighted exchangeability for covariate shift).

Planned extensions. Conformalized quantile regression (Romano et al., 2019) with self-calibration (van der Laan & Alaa, 2024) would learn heteroscedastic intervals that adapt to local uncertainty—tighter for well-characterized biomarkers and wider for novel ones. Localized conformal prediction (Guan, 2021) could further provide approximate conditional coverage in biomarker subgroups, though formal conditional guarantees require larger calibration sets than are currently available.

C VARIABLE DEFINITIONS AS INSTANTIATED

Table 1: Formal variables as instantiated in the illustrative analysis.

Symbol	Domain	Instantiation
X^{m}	Mouse features	Log_2 fold-changes of 174 DEPs (5xFAD, PS19)
Y^{m}	Mouse labels	Treatment response (reversal score per biomarker)
X^{h}	Human features	Matched ortholog expression from AD cohorts
Y^{h}	Human target	Published biomarker measurements (not treatment response)

Ortholog mapping. Mouse-to-human gene mapping uses NCBI HomoloGene (build 68). Of the 174 DEPs from Chaubey et al. (2026), 156 have one-to-one human orthologs. Features are harmonized by batch correction (ComBat) on shared pathway-level scores rather than individual gene expression.

Density-ratio estimation. The likelihood ratios $w_i = p_{\mathrm{h}}(X_i^{\mathrm{m}})/p_{\mathrm{m}}(X_i^{\mathrm{m}})$ are estimated via a logistic regression domain classifier trained to distinguish mouse from human samples using the harmonized feature space. Predicted probabilities are clipped to $[0.05, 0.95]$ to avoid extreme weights.

Interval width. At $\alpha = 0.10$, the weighted CP intervals have a mean width of 1.8 $\log_2$ fold-change units across the 12 biomarkers (range: 0.9–2.4). Concordant biomarkers tend to have narrower intervals (mean 1.5) than discordant ones (mean 2.3), reflecting higher predictive confidence for well-conserved biomarkers.

Data splits. Mouse proteomic data are split 60/20/20 for training, conformal calibration, and density-ratio estimation, respectively. The 12 human biomarkers serve as the test set. No human treatment-response data enter any training or calibration step.

Biomarker panel. The 12 representative biomarkers were selected a priori based on literature prevalence and data availability, before examining concordance results. They span: BBB markers (ZO-1, Claudin-5, CD13), inflammatory cytokines (IL-6, IL-1β), NAD^+ pathway enzymes (NAMPT, SIRT1), neurodegeneration markers (NfL, GFAP), synaptic markers (PSD-95, Synaptophysin), and the translatable plasma biomarker p-tau217.