Adjusting for Unmeasured Confounders in scRNA-seq Differential Analyses Using sensGAN with Application to Alzheimer's Disease

Yifan Lin[1], Taek Son[1], Kevin Z Lin[1]

University of Washington
Seattle, WA 98107, USA
yifanlin@uw.edu, sst91@uw.edu, kzlin@uw.edu

Abstract. Single-cell transcriptomic studies of Alzheimer's disease (AD) are confounded by unmeasured co-pathologies that can obscure disease-specific signals. We introduce *sensGAN*, a generative-adversarial sensitivity analysis framework that learns worst-case latent confounders and quantifies their impact on gene-level differential expression. By constraining the predictive gains of latent confounders on both disease status and gene expression, sensGAN identifies the weakest confounding strength required to nullify each gene's association with disease. Through simulations, we show that sensGAN accurately recovers latent confounding structure and outperforms existing surrogate variable and confounder-adjustment methods in identifying confounder-sensitive genes. In a proof-of-principle analysis of systemic lupus erythematosus (SLE), sensGAN identifies a subset of differentially expressed genes that are robust to increasing levels of latent confounding and learns confounders that align with held-out biological covariates. Applying sensGAN to microglia from the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD), we distinguish pathways that are robustly associated with AD pathology from those likely driven by co-occurring disease processes. By ranking genes according to their robustness to hidden confounding, sensGAN refines single-cell discoveries into comorbidity-aware, testable hypotheses and provides a principled framework for sensitivity analysis in high-dimensional genomics.

Keywords: single-cell RNA-seq, Alzheimer's disease, unmeasured confounding, sensitivity analysis, generative adversarial networks, differential expression, microglia, latent variable models

1 Introduction

Alzheimer's disease (AD) remains incurable, and disease-modifying therapies have so far produced only modest benefit. Microglia, the brain's resident immune cells, are a major therapeutic focus because they sense injury, remodel synapses, and clear misfolded proteins such as amyloid and tau. However, human single-cell studies of microglia are difficult to interpret: older individuals frequently harbor multiple neurodegenerative and systemic pathologies, so transcriptional changes attributed to AD may in fact reflect unmeasured comorbid processes. Similar challenges arise in single-cell studies of other complex disorders, from Parkinson's disease and frontotemporal dementia to autoimmune and cardiovascular disease, where environmental exposures and comorbid diagnoses are incompletely recorded.

Here we focus on microglial states across the AD continuum, but aim for a framework that generalizes to such multi-cause disease settings. We argue that, alongside conventional differential expression statistics, analyses should report how strong an unmeasured factor would need to be to fully explain each gene's association with the disease label. We introduce sensGAN, a generative-adversarial sensitivity analysis that learns worst-case latent confounders and quantifies their impact on gene-level significance. By ranking genes according to their robustness to hidden drivers, sensGAN turns single-cell results into comorbidity-aware, testable hypotheses for AD and other complex diseases.

1.1 Scientific relevance in AD research

One of the major computational obstacles of studying cellular changes in the brain during AD progression lies in the fact that when post-mortem brain tissue is sequenced, often the donor had co-occurring neurodegenerative diseases beyond AD [24, 19]. These co-occurring neurodegenerative diseases (i.e., "co-pathologies") potentially confound existing single-cell analyses. For example, in late-onset AD, only 31% of cases are described with AD-specific signatures [28]. To illustrate, we perform a preliminary analysis on a cohort of 4,168 deceased donors from the National Alzheimer's Coordinating Center (NACC), all of which were staged for AD as well as 5 other co-pathologies (Fig. 1A). As the severity of AD increases, an increasing percentage of those donors have more co-pathologies. These co-pathologies have a potential clinical impact, shortening the life-span of donors, even when adjusting for AD severity (Fig. 1B).

Although many of the cited studies, including our analysis shown in Fig. 1 on the NACC cohort, suggest that adjusting for co-pathologies is important to uncover the "pure" AD transcriptomic changes, the difficulty lies in the evolving definitions of these co-pathologies [11]. For instance Figure 1C summaries how when some of the major currently used clinical co-pathology staging criterias were standardized for AD research [20, 1, 22, 23]. There are major updates in our clinical diagnosis of brain tissue roughly every 10 years as our scientific understanding of this disease evolves. Given this ever-changing landscape, how do we ensure that our single-cell findings today are robust to possible confounding co-pathologies that we might formalize in the future?

Conceptually, we mirror Cornfield's classic sensitivity analysis of smoking and lung cancer, which argued that even without experimental evidence, no realistic unmeasured confounder could account for the observed association unless it were both extremely imbalanced between smokers and non-smokers and had an implausibly large effect on lung cancer risk [3]. Analogously, sensGAN quantifies how strong an unmeasured co-pathology would need to be, relative to observed covariates, to explain away each gene's association with AD, thereby distinguishing transcripts whose links to AD are robust to hidden disease processes from those that could plausibly be driven by currently unrecorded co-morbidities.

The purpose of this paper is to develop a framework for estimating the worst-case confounder while constraining predictive gains for the treatment (case-control) variable and high-dimensional results (gene expression). In the AD context, we are interested in how AD pathology ("treatment") impacts a cell's function, measured through gene expression ("outcome"). We use the term "confounder" to refer to a broad category of unmeasured covariates such as co-occurring neurodegenerative diseases, that have profound impact on the cellular responses to AD pathology.

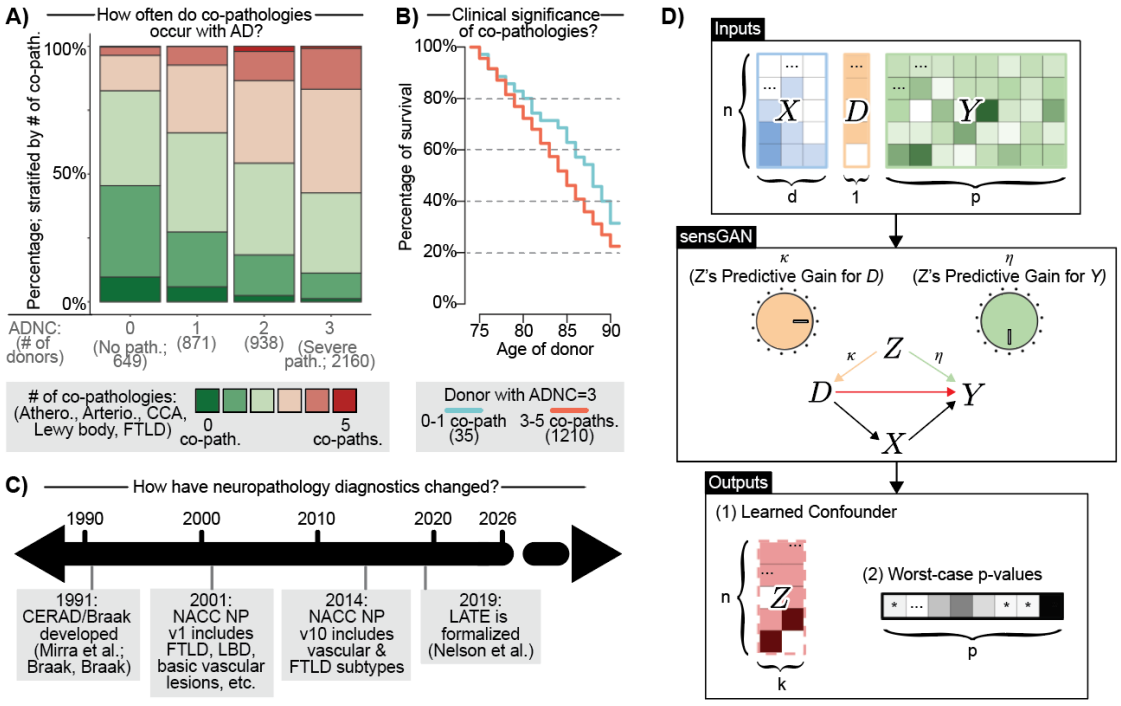

Fig. 1. Clinical impact of unmeasured confounders in AD research and sensGAN overview to learn them. A) Percentage of donors with number of co-pathologies, stratified by AD severity (ADNC score: 0 being no pathologies, 3 being severe). The number of donors in the NACC cohort in each stratum is marked. **B)** Survival curve among donors above the age of 75 with severe AD pathologies, stratified by whether they have 0/1 co-pathologies or 3+ co-pathologies. **C)** Timeline of when AD and related co-pathologies were standardized among clinical diagnostics of post-mortem brain tissue. **D)** Schematic sensGAN, showing that based on the donor covariates X, case-control D and pseudo-bulk RNA-seq profiles Y, sensGAN learns the worst-case confounders Z and confounder-adjusted p-values based two "knobs" which control how extreme the confounder is.

1.2 Existing computational methods and their limitations

Confounder adjustment with high-dimensional outcomes has been an important topic in statistics and genomics in recent years. To characterize the confounding effects, many works assume the prototypical model

$$Y \sim DT + XB + Z\Gamma + E,$$

where $Y \in \mathbb{R}^{n\times p}$ is the gene expression matrix across n donors and p genes, and each entry of Y is a random variable drawn from a distribution with mean defined by the $D \in \mathbb{R}^{n\times 1}$ (the measured case-control vector), $T \in \mathbb{R}^{1\times p}$ (the direct effect to be estimated), $X \in \mathbb{R}^{n\times d}$ (the observed covariate matrix for d measured confounders), $B \in \mathbb{R}^{d\times p}$ (the additional effects to be estimated), $Z \in \mathbb{R}^{n\times k}$ (the latent factor matrix for k unmeasured confounders), and $\Gamma \in \mathbb{R}^{k\times p}$ (the latent factor loading). However, since Z is unobserved, we are forced to model the data Y using a restricted model,

$$Y \sim D\widehat{T} + X\widehat{B}.$$

We highlight two computational frameworks that address unmeasured confounders Z in this setting that motivates our proposed method.

One framework investigates the coefficient B's maximum bias incurred by a hypothetical unmeasured confounder. This is called *omitted variable bias* (OVB), where the sensitivity analysis addresses how "extreme" an unmeasured confounder would need to be to nullify the significance of a treatment-outcome relation [2, 34]. For instance, sensmaker [2] proves for linear models, the OVB is reparameterized with partial R^2 for the case-control and outcome variable. However, this method is not easily generalizable to a non-linear model

and do not handle multiple outcomes. For instance, in our modeling of scRNA-seq data, we are interested in how the AD pathology impacts the thousands of genes (i.e., outcomes) via a negative-binomial (NB) generalized linear model (GLM) framework.

The other framework assumes that residual variation can be effectively captured by low-dimensional surrogate variables [13, 14]. This framework have been extended to methods such as CATE [36] and GCATE [8, 7] to model modern genomic data in the presence unmeasured confounders, i.e., thousands of genes modeled via the negative binomial GLM. While these methods provide an explicit estimate of the unmeasured confounders that can be used in any downstream analyses, they do not provide insight into the spectrum of confounders at varying levels of hypothetical confounding. Additionally, the estimated confounders are generally only reflective of residual variability in the genes (i.e., outcome) not reflected by the treatment and are independent of the treatment itself. In our context, this means these confounders are unlikely to reflect an unmeasured co-pathology that often co-occurs with AD.

2 Overview of sensGAN

Our method, **sensGAN** (Sensitivity Generative Adversarial Network), is an estimation framework based on GAN that (1) accommodates general relationships among observed covariates, multiple outcomes, and unmeasured confounders, (2) provides the worst-case confounders and quantifies their confounding effects, and (3) leveraging a deep-learning framework to efficiently search for the worst-case confounders for a spectrum of confounding effects. In our paper, we formalize "worst-case confounder" as the confounder that nullifies the largest number of significant genes. We explain each aspect separately below.

2.1 Quantifying impact of unmeasured confounder with predictive gain.

One of the most important ingredients of our framework is a quantification of how an unmeasured confounder Z relates to the case-control variable D and the outcome Y. We first define baseline prediction models f_D and f_Y as with generalized linear models (GLMs) that do not use any unmeasured confounders to predict D and Y, respectively.

$$\begin{aligned} &D \sim \text{Bernoulli}\big(f_D(X)\big), \quad \text{where} \quad f_D(X) = \text{logit}(D) = X\widehat{B}_{\text{res},D}, \quad \text{and} \\ &Y \sim \text{NB}\big(f_Y(X,D),\, \widehat{\alpha}_{\text{res}}\big) \quad \text{where} \quad f_Y(X,D) = \log(\mu) = D\widehat{T}_{\text{res},Y} + X\widehat{B}_{\text{res},Y}. \end{aligned}$$

To formulate how much an hypothetical matrix of unmeasured confounders $Z \in \mathbb{R}^{n\times k}$ improve the prediction of D and Y over the baseline models, we define adjusted models $\widetilde{f}_D$ and $\widetilde{f}_Y$.

$$D \sim \text{Bernoulli}\big(\widetilde{f}_D(X,Z)\big), \quad \text{where} \quad \widetilde{f}_D(X,Z) = \text{logit}(D) = X\widetilde{B}_D + Z\widetilde{G}_D, \quad \text{and} \tag{1}$$

$$Y \sim \text{NB}\big(\widetilde{f}_Y(X,D,Z),\, \widehat{\alpha}_{\text{res}}\big) \quad \text{where} \quad \widetilde{f}_Y(X,D,Z) = \log(\mu) = D\widetilde{T}_Y + X\widetilde{B}_Y + Z\widetilde{G}_Y. \tag{2}$$

We use deviance-based partial R^2, denoted as R^2_{Dev}, to quantify the difference in prediction performance between the model adjusting for the arbitrary confounder Z and the baseline model. The deviance of a model is defined by the log-likelihoods of the saturated model and the model of interest.

$$R^2_{\text{Dev},D} = \frac{\text{Dev}_{f_D} - \text{Dev}_{\widetilde{f}_D}}{\text{Dev}_{f_D}}, \quad R^2_{\text{Dev},Y} = \frac{\text{Dev}_{f_Y} - \text{Dev}_{\widetilde{f}_Y}}{\text{Dev}_{f_Y}}, \text{ where } \text{Dev}_f = 2\left[\ell(\text{saturated}) - \ell(f)\right] \tag{3}$$

f is any function for D and Y, including $f_D, f_Y, \tilde{f}_D, \tilde{f}_Y$ mentioned previously. Since R^2_{Dev} is not bounded, therefore difficult to interpret. To develop a bounded, interpretable metric, we develop an upper bound for predictive gains. Suppose there is the most powerful $\widehat{Z} \in \mathbb{R}^{n\times k}$. It is defined as the most powerful unmeasured confounder, which explains away possibly the most residual in D and Y unexplained by measured covariates. (We will describe how to compute $\widehat{Z}$ later in this section.) Using $\hat{Z}$ in (1) and (2) yields deviance-based partial R^2's, denoted as $\hat{R}^2_{\text{Dev},D}$ and $\hat{R}^2_{\text{Dev},Y}$, respectively, using (3). Equipped with these deviance-based partial R^2's, we now define interpretable metrics, normalized predictive gains κ and η,

$$\kappa = \widetilde{R}^2_{\text{Dev},D}/\widehat{R}^2_{\text{Dev},D} \in [0,1], \quad \text{and} \quad \eta = \widetilde{R}^2_{\text{Dev},Y}/\widehat{R}^2_{\text{Dev},Y} \in [0,1]. \tag{4}$$

These emulate similar properties and interpretations of partial R^2 for linear models as discussed in sensmarker [2]. The normalized predictive gain denotes the proportional predictive gain introduced by an arbitrary $\widetilde{Z}$, normalized by the maximal predictive gain introduced by the most powerful $\widehat{Z}$. A normalized predictive gain of 1 means that $\widetilde{Z}$ achieves 100% of the maximal predictive gain. Importantly, it is also differentiable, allowing for gradient calculation in a deep-learning model, as we describe later in this section.

2.2 Quantifying the "worst-case" p-value for a desired amount of predictive gain

After identifying the impact of confounders with normalized predictive gains, we explore the "worst-case" confounder among all confounders with certain predictive gains. Operationally, this is the confounder that nullifies the largest number of differentially expressed genes (DEGs). Given desired predictive gains κ' and η', we formulate estimating such a worst-case confounder as an optimization under constraints,

$$\max_{\widetilde{Z}} \left\{ \sum_{j=1}^{p} \mathbb{I}\left(\frac{[\widetilde{B}_Y]_j}{\mathrm{SE}([\widehat{B}_{\mathrm{res},Y}]_j)} - z_{1-\alpha} > 0 \right) \right\}, \quad \text{subject to} \quad \kappa = \kappa', \quad \eta = \eta', \tag{5}$$

where $z_{1-\alpha}$ is the $(1-\alpha)\times 100\%$ quantile of a Gaussian distribution, and $\widetilde{B}_j$ and B_j defined as the coefficient for the treatment D when adjusted and not adjusted for the unmeasured confounder, respectively. To compute the p-value of gene j adjusting for Z, we assume that $\mathrm{SE}(\widehat{B}_j) \approx \mathrm{SE}(\widetilde{B}_j)$, so we can model the DE test statistic $|\widetilde{B}_j/\mathrm{SE}(\widehat{B}_j)|$ as a z-score.

2.3 Deploying a deep-learning framework for efficient training of a spectrum of unmeasured confounders

To tie all the different components of sensGAN together, we adopt a neural network (NN) systems, sensGAN, to solve the optimization to identify an unmeasured confounder. The predictor consists of shallow fully connected networks that serve as differentiable surrogates for logistic and NB GLMs. Given the inputs, the predictor outputs the logit for D and the NB mean for Y. The overdispersion parameters are estimated from the baseline NB GLM. Because each predictor is a single linear layer, maximizing its likelihood is equivalent to solving for the maximum likelihood estimators of the corresponding GLMs. The generator is a two-layer fully connected network that learns a continuous latent confounder Z from X, D, and Y. It maps inputs into a hidden representation and then outputs a mean vector $\mu(X)$ and a log-variance vector $\log \sigma^2(X)$, from which a latent sample is drawn using a reparameterization-style step. The sampled latent vector is finally passed through a sigmoid transformation to yield $Z \in (0, 1)$.

Motivated by Equation (5), sensGAN poses an optimization problem that eliminates the significance of the largest number of features (genes) associated with D while maintaining plausibility by limiting the predictive gains that accounted for the confounding. Here, sensGAN has 3 steps: (1) train a NN to model the baseline predictors; (2) train a NN to estimate the most powerful confounder $\widehat{Z}$; (3) train a the multitask GAN to generate the worst-case $\widetilde{Z}$ given normalized predictive gain constraints, yielding the worst-case significance profile under constraints. We describe each step below:

Baseline predictors. In the first step, we train the baseline predictor NNs with the objective to maximize likelihood. This step takes in X to predict D, and X, D to predict Y.

$$\text{D Predictor Objective} = \lambda_{\mathrm{BCE}} \frac{1}{n} \sum_{i=1}^{n} \ell_i^{\mathrm{BCE}}, \quad \text{Y Predictor Objective} = \lambda_{\mathrm{NB}} \frac{1}{n} \sum_{i=1}^{n} \ell_i^{\mathrm{NB}}$$

Most powerful confounder. In the second step, we train a NN system to identify the most powerful confounder, $\widehat{Z}$, and the corresponding predictors that additionally took $\widehat{Z}$ as one of the inputs. The NN system consists of 2 components: (1) The predictors take in X, $\widehat{Z}$ to predict D, and X, $\widehat{Z}$, D to predict Y. Their objectives are the same as those of baseline predictors. (2) The generator takes in X, D, and Y to generate the $\widehat{Z}$ that minimized the likelihood losses:

$$\text{Generator Objective} = D \text{ Predictor Objective} + Y \text{ Predictor Objective}$$

Then, we calculated the maximal predictive gains introduced by the most powerful confounder $\widehat{Z}$, $\widehat{R}^2_{D,\text{Dev}}$, and $\widehat{R}^2_{Y,\text{Dev}}$.

Plausible confounders constrained by predictive gains. The most powerful confounders explain as much residual in D and Y as possible, which is typically unlikely to be a real clinical measurement like co-pathology. Therefore, we propose the plausible confounder $\widetilde{Z}$, whose predictive gains are bounded by those introduced by the most powerful confounder $\widehat{Z}$. Following the same line of calculation in Equation 3, we calculated the predictive gains $\widetilde{R}^2_{\text{Dev, D}}$ and $\widetilde{R}^2_{\text{Dev, Y}}$. With the maximal predictive gains calculated in the previous step, we have the normalized predictive gains of this plausible confounder, $(\kappa',\, \eta') \in [0,1]^2$.

Next, we utilized the GAN system to identify the unmeasured confounder with certain predictive gains $(\kappa',\, \eta')$ and the worst-case p-value; we adopt the GAN system (Fig. 2A). The architecture of the GAN system is similar to the NN system in step 2, but the predictor and the generator now have adversarial objectives. It's because the predictors maximize likelihoods while the generator controls for predictive gains, which is positively proportional to likelihoods as implied in Equation 3.

$$\text{GAN Generator Objective} = \min \left\{ \lambda_{\text{pval}} \cdot \frac{1}{p} \sum_{j=1}^{p} \sigma_\lambda \left(\left| \frac{\widetilde{B}_{Y,j}}{\text{SE}(\widehat{B}_{\text{res},Y,j})} \right| - z_{1-\alpha} \right) + \lambda_Y \left[\left((\eta - \eta')\, s_Y \right)^2 \right] + \lambda_D \left[\left((\kappa - \kappa')\, s_D \right)^2 \right] \right\}$$

To compute the DE p-value of gene j adjusting for $\widetilde{Z}$, we assume that $\text{SE}(\widehat{B}_{j,\text{res}}) \approx \text{SE}(\widehat{B}_j)$, so we can model the DE test statistic $|\widehat{B}_j / \underline{\text{SE}}(\widehat{B}_{j,\text{res}})|$ as a z-score. Subsequently, we identify the worst-case p-values for all features, a p-value vector $\widetilde{P} \in \mathbb{R}^{p \times 1}$.

We initialize the GAN by solving a ridge-regularized linear regression that maps the generator's hidden layer to the logit-transformed $\widehat{Z}$. This yields closed-form estimates for the weights and bias of the μ-head, ensuring that the generator starts from $\mu(X) \approx \text{logit}(\widehat{Z})$ before the alternating training begins. At the end of training, the GAN finds the confounder, $\widetilde{Z}$, which resulted in the worst-case gene significance profile while maintaining certain predictive gains.

Sequential training of plausible confounders constrained by a spectrum of predictive gains. To span over all plausible confounders with non-negative predictive gains, we trained the sensGAN model sequentially with a sweep-through of the predictive gains of (κ', η') over the unit interval $[0, 1]$. At the end of the sequential training, the generator found a series of worst-case confounders with varied predictive gains.

2.4 Downstream analysis: Calibration of predictive gains with measured covariates

From previous results, we have a series of plausible confounders with the worst-case p-values characterized by the normalized predictive gains $(\kappa',\, \eta')$, w.r.t. the case-control variable D and the gene counts Y. How do we interpret them relative to measured covariates?

The calibration step calculates the predictive gains of measured covariates so that unmeasured confounders and measured covariates are comparable. Suppose the strength of the confounder is x times the strength of the most powerful measured covariate, and sensGAN learned that this confounder, in the worst-case, nullifies half of the DEGs.

If $x = 10$, it means the confounder is 10 times as powerful as the most powerful measured covariate, which is very unlikely to be a true clinical measurement that is missed. In this case, we conclude that significant DEGs are nullified under very strong unmeasured confounding, therefore more robust.

If $x = 0.1$, it means that the most powerful measured covariate is 10 times as powerful as the confounder, which is very likely a true clinical measurement. In this case, we conclude that significant DEGs are nullified under very weak unmeasured confounding, therefore less robust.

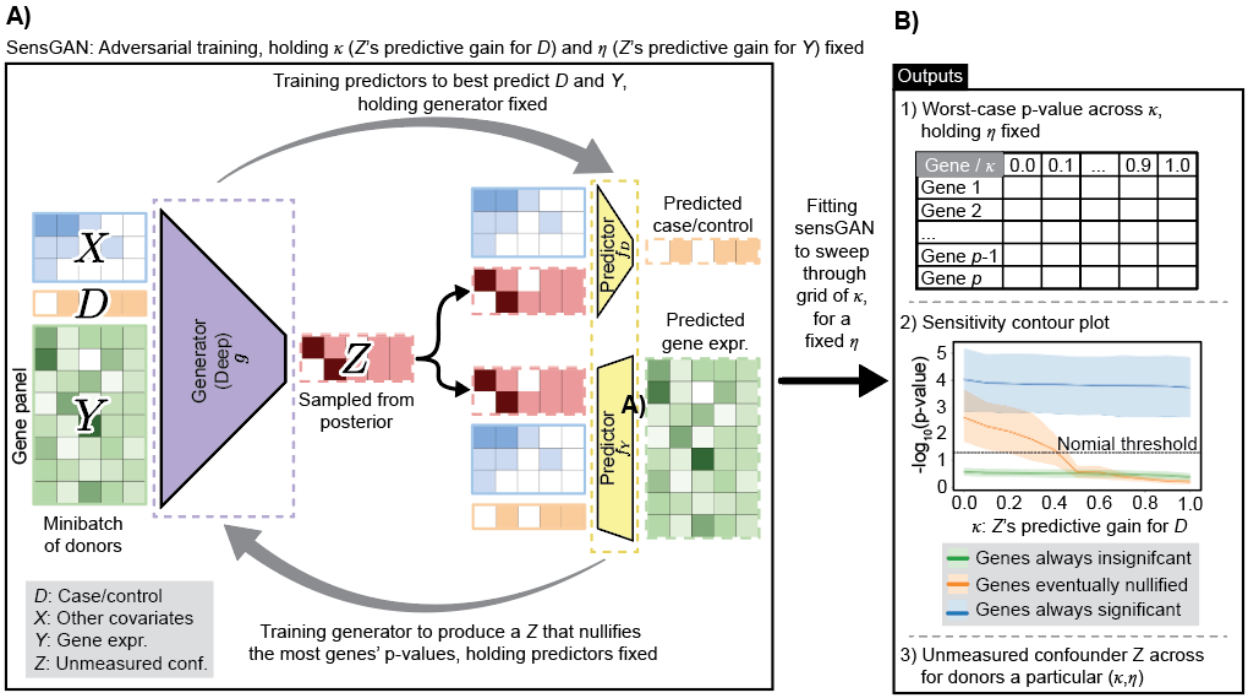

Fig. 2. sensGAN method overview. A) The GAN system that learns the confounder Z with specific predictive gains (κ, η). **B)** Method outputs while sweeping through a grid of κ and holding η fixed.

2.5 Downstream analysis: diagnostic sensitivity contour plots

Next, we provide a diagnostic tool, the "extreme-scenario" sensitivity plot originally proposed in [2]. Holding η fixed, we allow κ to vary and document the corresponding worst-case p-values. For every gene, we first draw the coordinates (κ', worst-case p-value) based on the output of sensGAN sweepthrough analyses. Then, we connect the coordinates and smooth the curve by an isotonic regression to make sure that the worst-case p-values could only stay the same or become more inflated as the confounder gets stronger. One can then see how strong such confounders would need to be w.r.t. D for the significant DEGs to be nullified. If a gene loses significance at a small κ, it is more likely to be associated with a confounder instead of the case-control variable. If it remains significant until a large κ, it is more likely to respond directly to the case-control variable.

3 Experimental Setup

3.1 Analysis of the simulation dataset for method comparison

We evaluated the performance of our approach using a simulated dataset. First, we simulated the measured covariate matrix $X \in \mathbb{R}^{n \times d}$ and the unmeasured binary confounder $Z \in \mathbb{R}^{n \times k}$ independently. Then, we simulated the case-control variable $D \in \mathbb{R}^{n \times 1}$ with a Bernoulli distribution with a logit link while controlling for the variance contributed by terms XB_1 and ZG_1. With the Negative Binomial GLM link function, we simulated the mean count matrix of scRNA-seq pseudocounts $Y \in \mathbb{R}^{n \times p}$. Then, we simulated the overdispersion rate to be a normal distribution centered around a constant vector $C \in \mathbb{R}^{p \times 1}$.

$$D \sim \text{Bernoulli}(\sigma(XB_1 + ZG_1)),$$

$$Y \sim \text{NegBinomial}(\log(\mu), \alpha), \text{where} \quad \log(\mu) = (X, D)\, B_2 + ZG_2, \alpha = C$$

With different combinations of $B_2 \in \mathbb{R}^{(d+1) \times p}$ and $G_2 \in \mathbb{R}^{k \times p}$, we simulated 3 categories of genes with varied significance profiles:

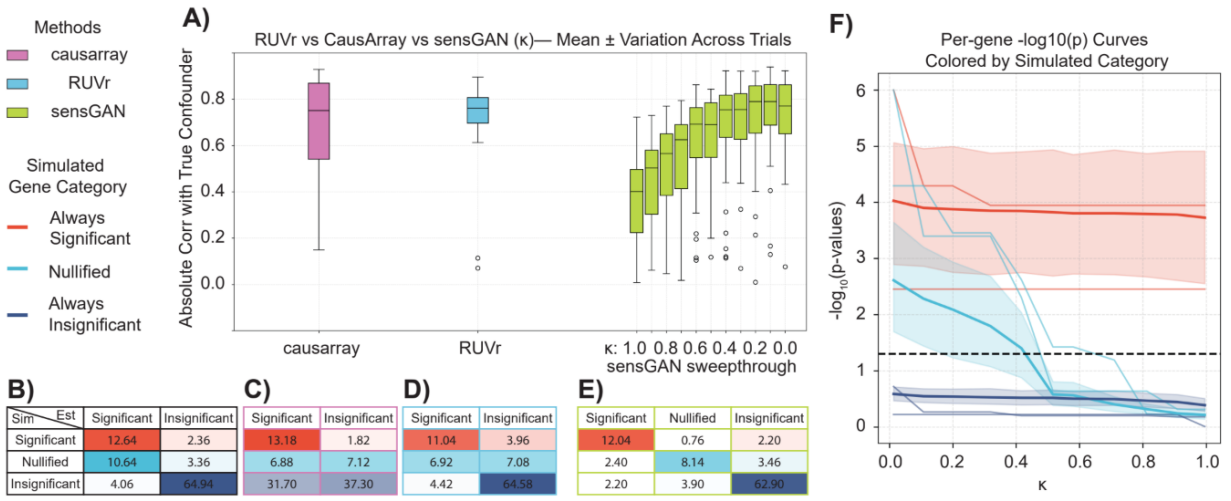

Fig. 3. Simulation demonstrates sensGAN's accuracy over other methods A) Correlations between the true confounder and the confounders estimated by causarray, RUVr, and sensGAN. **B)** The contingency table of gene significance between the simulation and the estimation by the naive GLM negative binomial regression model. **C)** The contingency table of gene significance between the simulation and the estimation by the causarray. **D)** The contingency table of gene significance between the simulation and the estimation by RUVr. **E)** The contingency table of gene significance between the simulation and the estimation by sensGAN. **F)** Sensitivity diagnostic contour plot with fixed $\eta = 0.5$. The thick curve represents the mean across genes, the thin curve shows an individual gene for illustration, and the shaded band denotes the 95% confidence interval.

Category 1: Genes significantly associated with D but not with Z. Their expression remains significantly associated with D both with and without adjusting for Z. They are later referred to as the (always) significant genes.

Category 2: Genes significantly associated with Z but not with D. These genes appear significant without adjusting for Z but become insignificant once Z is included in the model. They are later referred to as the nullified genes.

Category 3: Genes not significantly associated with either D or Z. They remain insignificant with and without adjustment for Z. They are later referred to as the (always) insignificant genes.

In our simulation, we set $n = 100, p = 100, d = 4, k = 1, C = \{0.5\}^p$. Three methods are applied to the simulation datasets: (1) GCATE, the unified statistical estimation and inference framework with the Poisson likelihood. (2) RUVr, the statistical method that removes unwanted residuals with high-dimensional data. (3) sensGAN, our proposed method that exploits the theory of sensitivity analysis and produces a series of potential confounders with varying predictive gains. Here, we fixed the predictive gain knob for outcome (Y) at 0.5 while turning the predictive gain knob for the case-control variable (D).

To evaluate different methods, we first compared the correlations between the learned confounder and the true confounder (Fig. 3A). The most powerful confounder identified by sensGAN achieves a mean correlation comparable to that obtained by causarray and RUVr. As the predictive gain knob, κ, is turned from 1 to 0, the mean correlation between the learned and true confounders decreases smoothly from approximately 0.75 to 0.4. sensGAN not only identifies the strongest latent confounding structure—on par with competing approaches—but also generates a continuum of plausible worst-case confounders constrained by predictive gains.

We then compare the contingency tables of gene significance across methods. The baseline GLM method, causarray, RUVr, and sensGAN are all capable of identifying the always significant genes (Fig. 3B, C, D, E). causarray (Fig. 3D) behaves conservatively while identifying the always insignificant genes, while others are capable of identifying genes in this category. The compromised causarray performance is likely because the method is designed for a single-cell resolution instead of a donor-level pseudo-bulk resolution.

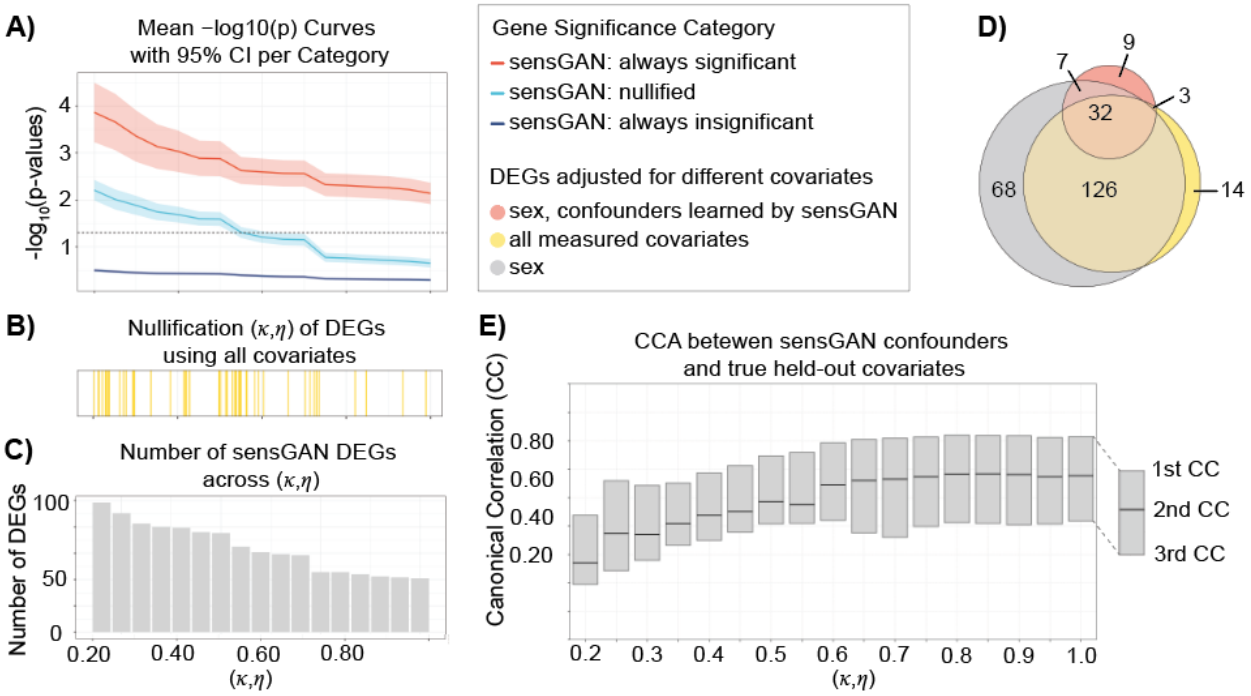

Fig. 4. sensGAN identifies robust SLE DEGs under increasing latent confounding and learns biologically meaningful confounders aligned with held-out covariates. A) The sensitivity diagnostic contour plot with (κ, η), where $\kappa = \eta$, varying from 0.2 to 1. **B)** The κ and η values where genes are nullified, extrapolated from the stepwise curve. **C)** The barplot of the number of DEGs at varying (κ, η) pairs. **D)** The Venn plot of significant DEGs calculated by the GLM Gamma-Poisson model adjusting for different covariates. **E)** The canonical correlations between the sensGAN confounders at varying levels of strength and the held-out covariates.

With the nullified genes, the baseline GLM method naively considers them significant (Fig. 3B), while causarray and RUVr consider half of them significant and the other half insignificant (Fig. 3C, D). In contrasts, sensGAN learns a series of potential confounders with varying predictive gains, on average capable of identifying 54.27% of the nullified genes (Fig. 3E). However, we don't expect sensGAN to identify all the nullified genes. We expect sensGAN to consider some genes always significant or always insignificant because we fix one of the predictive gain knobs. At the same time, sensGAN maintains high accuracy in identifying the correct signal patterns for the always-significant genes and the always-insignificant genes.

We further present the sensitivity diagnostic contour plot based on sensGAN outputs(Fig. 3F). The y-axis is the logarithm of the worst-case p-values, the black dash line is the significance threshold of 0.05, and the x-axis is the predictive knob (κ) for the case-control variable. Every gene has a contour curve that depicts its sensitivity to the estimated unmeasured confounder, exemplified by the thin curves. The thick curves are the mean contour curves, and the bands are the 95% confidence intervals, colored by the simulated gene categories. If a gene crosses the significance threshold with a smaller κ, the gene is considered to be nullified by a weaker confounder. Compared to genes nullified by stronger confounders, this gene is less directly associated with the case-control variable. Therefore, the differential expression conclusion is less robust.

3.2 Analysis of systemic lupus erythematosus as proof-of-principle

Systemic lupus erythematosus (SLE) is an autoimmune disease predominantly affecting women and individuals of Asian, African, and Hispanic descent [26]. Multiplexed single-cell RNA sequencing (mux-seq) was developed to capture the complexity of immune cell populations and systematically profile the composition and transcriptional states of immune cells in a large multiethnic cohort. The dataset contains 1.2 million peripheral blood mononuclear cells from 8 major cell types and 261 individuals, including 162 SLE cases and

99 healthy controls of either Asian or European ancestry. The cell-type-specific DE analysis aims to provide insights into the diagnosis and treatment of SLE.

To remove the genes with small variations, we pre-process the single-cell data by selecting the top 2,000 highly variable genes (HVGs) within each cell type, computing the pseudo-bulk profiles for each donor, and log-normalized counts based on the total library size per cell. With a focus on the T4 cell type, we aggregated single-cell expression profiles by summing counts across cells from the same subject, yielding a gene-level pseudo-bulk count matrix. Then, we removed genes with over 90% zero counts across subjects, retaining those expressed in at least 10% of samples for downstream analysis. For the stablity of method performance, we also removed the genes whose counts didn't converge in the naiive GLM Negative Binomial Model. In total, the dataset included 256 subjects (158 cases and 98 controls). For each subject, the case–control variable represented systemic lupus erythematosus (SLE) status and the final number of retained genes was 475.

In the proof-of-principle analysis, we only provide the method with sex while holding out other measured covariates, including the population covariate and the processing cohort information ($d = 5$, $d' = 4$). To make sure the method captures biologically meaningful confounders, we decided to search for confounders ($k = 3$) with a wider range of confounding effects by letting κ and η be equal to each other and vary together, i.e. letting Z's predictive gains on Y be equal to Z's predictive gains on X. With a series of (κ, η) from 0.2 to 1 with a stepsize of 0.05, the method identified a series of confounders. We then performed adjusted differential expression analyses with confounders of varying strengths.

First, we analyzed the significance profiles of genes under varying levels of confounder strengths. In the sensitivity contour plot (Fig. 4A), mean contour curves and 95% confidence intervals were colored by genes that are always significant, nullified, and always insignificant. Then, we calculated the exact κ at which the gene became insignificant by extrapolating on the stepwise contour curve (Fig. 4B). Shown by the bar graph of the number of differentially expressed genes in the κs, we also observed that larger κ generally corresponded to fewer differentially expressed genes (Fig. 4C). These results showed that the method is capable of identifying a series of confounders with different strengths, and that more significant DEGs in the original model became insignificant as κ increased.

Then, we analyzed the DEG sets across 3 methods to understand how sensGAN compares to existing methods while applying to real data. We used the Gamma-Poisson regression model to identify the DEGs associated with SLE, adjusting for different sets of covariates. For the sensGAN method, we adjusted for sex and the sensGAN-learned confounders. For the all-covariate method, we adjusted for all measured covariates. For the baseline method, we only adjusted for sex. We presented the results as a Venn diagram, in which circle areas correspond to the total number of DEGs identified while adjusting for different sets of covariates (Fig. 4D). The baseline DEG set was the largest, because it only adjusted for sex. The all-covariate DEG set was smaller than the baseline DEG set and larger than the sensGAN DEG set, due to its adjustment of other measured covariates. The sensGAN DEG set only includes the genes that are always significant with varying levels of confounding strengths; therefore, the smallest set, most of which were also identified by the other methods.

Lastly, we summarized relationships between the confounders ($k = 3$) learned by sesnGAN and the heldout covariates ($d' = 4$) by canonical correlation analyses (CCA) (Fig. 4E). Every bar corresponds to 3 canonical correlation (CC) values while comparing the heldout covariates and the learned confounders. We observed that all CCs increased as the confounding strengths increased from 0.2 to 1. Through this observation, we believed that the method was capable of learning a series of biologically meaningful confounders, which became more and more similar to the held-out covariates as we allowed for stronger confounding effects.

3.3 Analysis of microglia studying Alzheimer's disease, adjusting for unmeasured co-pathology

After verifying the reliability of sensGAN on simulated and a well-studied biological system, we now turn our attention towards studying Alzheimer's disease (AD), our primary disease of interest. We leverage the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) consortium [10], a single-cell brain atlas of comprising of 80+ donors to study AD. We focus on the microglia in the prefrontal cortex (PFC) due to the microglia's critical role in clearing AD pathology and its implicated from established GWAS studies [12, 31]. We construct a pseudo-bulk profile, similar to our analysis in Section 3.2, and perform our sensGAN analysis asdjusting for sex, age, APOE4 status, ethnicity, and PMI. We focus on a key set of genes that had a p-value < 0.2

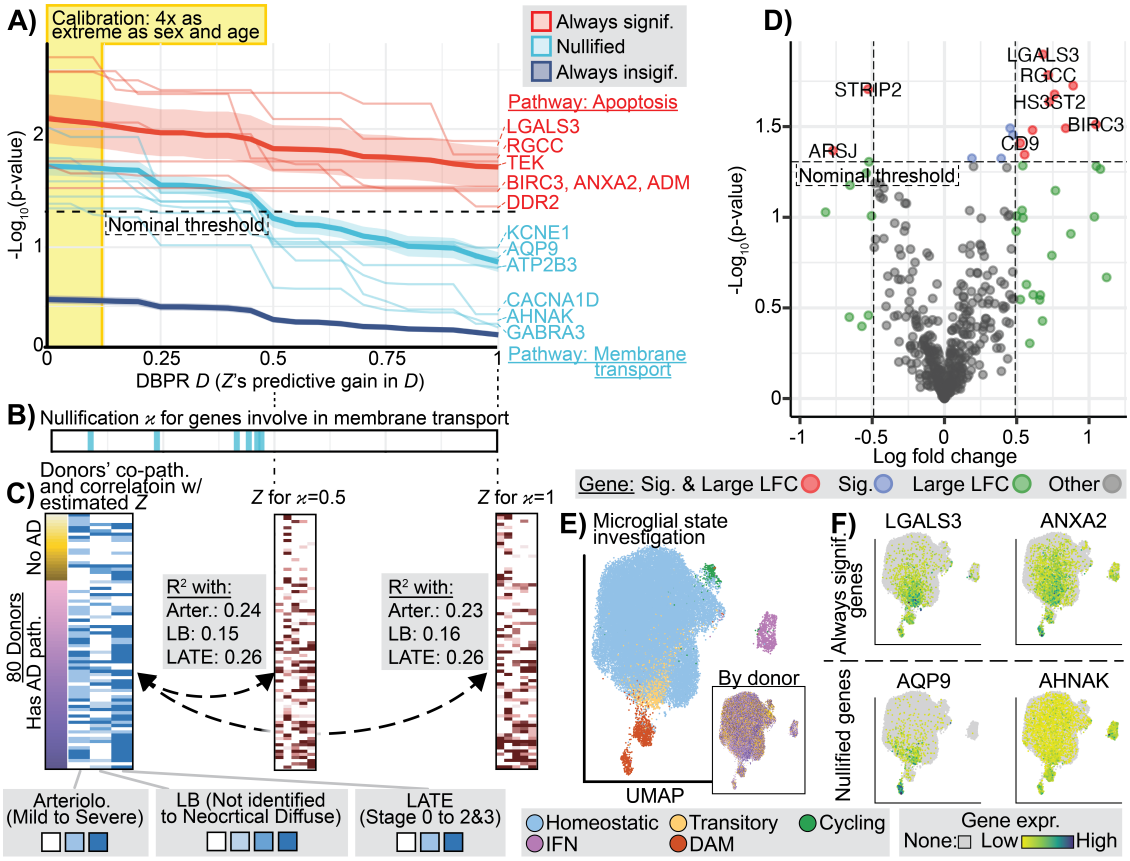

Fig. 5. sensGAN prioritizes DEGs specific to microglia in AD. A) Diagnostic contour, showing genes that are always significant (red), nullified (light blue), and always insignificant (dark blue) across different values of κ. The mean trend with 95% band is shown, and genes in enriched pathways are shown. **B)** The κ of when genes in the membrane transport pathway are nullified. **C)** Donors (rows) and their measured co-pathology, as well as the estimated unmeasured confounders for $\kappa \in \{0.5, 1\}$. The R^2 based on linear prediction for each co-pathology is shown. **D)** Volcano plot based on adjusting for the unmeasured confounder with $\kappa = 1$. **E)** UMAP of single-nuclei microglia in SEA-AD, colored by microglial state and donor. **F)** Expression of certain genes across the microglia.

when doing a differential expression test via DESeq2 [17], resulting in 448 genes in our analysis. While many of these genes in our analysis are already nominally significant, we ask whether some of the genes will be nullified by the unmeasured confounders discovered by sensGAN.

Our sensGAN analysis, sweeping across κ from 0 to 1, disentangles potential microglial pathways that are possibly confounded by co-pathologies from others that are specific to AD. Our diagnostic contour plots reveals that genes that are always nominally significant are enriched for an apoptosis pathway (red; Fig. 5A) [5]. For instance, the LGALS3 gene is strongly upregulated in plaque-associated microglia in AD models and human tissue [33], and has been implicated to be AD-specific through knockouts in AD mouse models, which demonstrated reduced plaque burden and improved cognitive behaviors [32]. In stark contrast, genes that are eventually nullified by sensGAN's learned unmeasured confounders Z are enriched for transmembrane transport (light blue). To calibrate the range of κ needed to nullify these genes, we also mark the range of κ up to four times as extreme as the predictive power of sex and age. For instance, AHNAK regulates voltage-gated calcium channels [18]. Although it is nominally significant without any adjustment for confounders, this gene is quickly nullified by sensGAN. This is plausible since its association with AD is not solely through microglia, but rather its interaction with neurons [35]. Furthermore, it has been suggested to be also involved in Lewy bodies [30] and Frontal Temporal Dementia [16], two other co-occurring pathologies typically associated with AD. The nullification κ for AHNAK and other genes in this pathway are shown in Figure 5B. Additionally, AQP9, a gene to regulate membrane channels to conduct water, is also nullified,

which is canonically associated with an AD-relevant gene for astrocytes, not microglia [15]. Together, these results demonstrate that sensGAN can meaningfully "purify" existing DEG results to retain only microglia-specific DEGs associated with AD (Fig. 5D for $\kappa = 1$).

Next, we illustrate that sensGAN's unmeasured confounders are suggestive of currently measured co-pathologies. While we did not use any co-pathologies in our sensGAN analysis, the SEA-AD consortium also staged all the post-mortem tissues for arteriolosclerosis (i.e., a vascular disease), Lewy bodies, and Limbic-predominant age-related TDP-43 encephalopathy (LATE). We compute how predictive sensGAN's learned confounders predict these co-pathologies (Fig. 5C). Surprisingly, for both $\kappa \in \{0.5, 1\}$, the predictive power measured by R^2 between the predicted (linear regression) and observed co-pathology stages is significantly non-zero and not changing. This suggests that even with a lower $\kappa < 0.5$, the sensGAN unmeasured confounders are already capturing all the co-pathology impacts from snRNA-seq data. We hypothesize that for $\kappa > 0.5$, the unmeasured confounders are capturing impacts beyond co-pathology, for instance impacts of spatial microenvironments, genetics, or environmental impacts (i.e., diet) on the donor that impact AD.

We lastly investigate which microglial states harbor the DEGs (both nullified or always significant). While our sensGAN analysis was performed on the pseudo-bulk level, the localization to specific microglial states reveals the underlying biological mechanism [27]. We first visualize the microglia in the SEA-AD cohort by the annotated microglial state (Fig. 5E). Then, we find that for both genes that were always significant regardless of κ (i.e., LGALS3 and ANXA2) or nullified genes (i.e., AQP9, AHNAK), the elevated gene expression typically occurs in DAM and transitory microglia (Fig. 5F). This demonstrates the biological insight uniquely uncovered by sensGAN – although many the originally implicated DEGs without any unmeasured confounders are localized to DAM and transitory microglia, sensGAN enables us to further prioritize these DEGs based on how "pure" they are for microglia in AD.

3.4 Discussion

We introduced sensGAN as a generative-adversarial sensitivity analysis framework for quantifying the vulnerability of single-cell differential expression results to unmeasured confounding. Rather than attempting to directly estimate the true confounders, sensGAN identifies worst-case latent confounders under explicit predictive-gain constraints, yielding interpretable sensitivity diagnostics for each gene. This design allows investigators to assess not only whether a gene is statistically significant, but also how robust that significance is to plausible hidden biological drivers.

Across simulations, sensGAN accurately recovered latent confounding structure and maintained high fidelity in separating always-significant, nullified, and always-insignificant genes, while also generating a continuum of plausible confounders with varying strengths. In contrast to existing surrogate-variable approaches, sensGAN explicitly models the spectrum of hypothetical confounding rather than committing to a single latent correction, enabling a principled robustness analysis.

In the systemic lupus erythematosus proof-of-principle study, sensGAN produced smaller and more stable DEG sets than baseline or fully adjusted models, while learning latent confounders that increasingly aligned with held-out biological covariates under stronger confounding. These results demonstrate that sensGAN can recover biologically meaningful latent structure despite being trained without access to those covariates.

In the SEA-AD microglia analysis, sensGAN further distinguished pathways likely specific to AD pathology from those plausibly driven by co-pathologies such as vascular disease or Lewy body pathology. Genes that remained robust under strong confounding were enriched for apoptosis-related processes, whereas genes that were rapidly nullified were enriched for transmembrane transport and were previously implicated in non-AD neurodegenerative conditions. This ability to prioritize disease-specific microglial programs addresses a central limitation of post-mortem single-cell studies, where multiple co-occurring pathologies are the norm rather than the exception.

Several limitations warrant future investigation. First, worst-case confounders learned by sensGAN are not guaranteed to correspond uniquely to any single biological process, and may conflate multiple latent sources. Second, the current framework relies on pseudo-bulk aggregation and generalized linear models, and extension to full cell-level hierarchical models remains an important direction. Finally, while predictive gain provides a calibrated measure of confounding strength relative to observed covariates, translating these quantities into absolute biological effect sizes remains an open challenge.

Overall, sensGAN provides a unified framework for adversarial sensitivity analysis in high-dimensional genomics. By shifting the focus from binary significance to quantitative robustness, sensGAN enables more reliable interpretation of single-cell disease mechanisms in the presence of pervasive and evolving confounding.

Bibliography

[1] Braak, H., and Braak, E. Neuropathological stageing of alzheimer-related changes. *Acta neuropathologica 82*, 4 (1991), 239–259.

[2] Cinelli, C., and Hazlett, C. Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology 82*, 1 (2020), 39–67.

[3] Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., and Wynder, E. L. Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer institute 22*, 1 (1959), 173–203.

[4] Di Mari, R., Ingrassia, S., and Punzo, A. Local and overall deviance r-squared measures for mixtures of generalized linear models. *Journal of classification 40*, 2 (2023), 233–266.

[5] Dou, R.-x., Zhang, Y.-m., Hu, X.-j., Gao, F.-L., Zhang, L.-L., Liang, Y.-h., Zhang, Y.-y., Yao, Y.-p., Yin, L., Zhang, Y., et al. Aβ1-42 promotes microglial activation and apoptosis in the progression of ad by binding to tlr4. *Redox biology 78* (2024), 103428.

[6] Drange, O. K., Smeland, O. B., Shadrin, A. A., Finseth, P. I., Witoelar, A., Frei, O., Group, P. G. C. B. D. W., Wang, Y., Hassani, S., Djurovic, S., et al. Genetic overlap between Alzheimer's disease and bipolar disorder implicates the MARK2 and VAC14 genes. *Frontiers in Neuroscience 13* (2019), 220.

[7] Du, J.-H., Shen, M., Mathys, H., and Roeder, K. Causal differential expression analysis under unmeasured confounders with causarray. *bioRxiv* (2025), 2025–01.

[8] Du, J.-H., Wasserman, L., and Roeder, K. Simultaneous inference for generalized linear models with unmeasured confounders. *Journal of the American Statistical Association*, just-accepted (2025), 1–24.

[9] Filippova, G. N., Casad, M., Groneck, C., Hui, K., Mishra, S., MacDonald, J. W., Bammler, T., Van Dyke, D. L., Skakkebaek, A., Gravholt, C. H., et al. Modeling sex differences in Alzheimer's disease using isogenic hiPSC lines with different sex chromosome complements and APOE alleles. *Alzheimer's & Dementia 20*, Suppl 1 (2025), e093548.

[10] Gabitto, M. I., Travaglini, K. J., Rachleff, V. M., Kaplan, E. S., Long, B., Ariza, J., Ding, Y., Mahoney, J. T., Dee, N., Goldy, J., et al. Integrated multimodal cell atlas of Alzheimer's disease. *Nature Neuroscience 27*, 12 (2024), 2366–2383.

[11] Jack Jr, C. R., Andrews, J. S., Beach, T. G., Buracchio, T., Dunn, B., Graf, A., Hansson, O., Ho, C., Jagust, W., McDade, E., et al. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimer's & Dementia 20*, 8 (2024), 5143–5169.

[12] Karch, C. M., and Goate, A. M. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biological psychiatry 77*, 1 (2015), 43–51.

[13] Leek, J. T., and Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS genetics 3*, 9 (2007), e161.

[14] Leek, J. T., and Storey, J. D. A general framework for multiple testing dependence. *Proceedings of the National Academy of Sciences 105*, 48 (2008), 18718–18723.

[15] Liu, J.-Y., Chen, X.-X., Chen, H.-Y., Shi, J., Leung, G. P.-H., Tang, S. C.-W., Lao, L.-X., Yip, H. K.-F., Lee, K.-F., Sze, S. C.-W., et al. Downregulation of aquaporin 9 exacerbates beta-amyloid-induced neurotoxicity in alzheimer's disease models in vitro and in vivo. *Neuroscience 394* (2018), 72–82.

[16] Lorenzini, I., Alsop, E., Levy, J., Gittings, L. M., Lall, D., Rabichow, B. E., Moore, S., Pevey, R., Bustos, L. M., Burciu, C., et al. Moderate intrinsic phenotypic alterations in c9orf72 als/ftd ipsc-microglia despite the presence of c9orf72 pathological features. *Frontiers in cellular neuroscience 17* (2023), 1179796.

[17] Love, M. I., Huber, W., and Anders, S. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology 15*, 12 (2014), 550.

[18] Matza, D., Badou, A., Kobayashi, K. S., Goldsmith-Pestana, K., Masuda, Y., Komuro, A., McMahon-Pratt, D., Marchesi, V. T., and Flavell, R. A. A scaffold protein, ahnak1, is required for calcium signaling during t cell activation. *Immunity 28*, 1 (2008), 64–74.

[19] McAleese, K. E., Colloby, S. J., Thomas, A. J., Al-Sarraj, S., Ansorge, O., Neal, J., Roncaroli, F., Love, S., Francis, P. T., and Attems, J. Concomitant neurodegenerative pathologies contribute to the transition from mild cognitive impairment to dementia. *Alzheimer's & Dementia 17*, 7 (2021), 1121–1133.

[20] Mirra, S. S., Heyman, A., McKeel, D., Sumi, S., Crain, B. J., Brownlee, L., Vogel, F., Hughes, J., Belle, G. v., Berg, L., et al. The consortium to establish a registry for alzheimer's disease (cerad) part ii. standardization of the neuropathologic assessment of alzheimer's disease. *Neurology 41*, 4 (1991), 479–479.

[21] Mittlböck, T. W. M., Haidinger, G., Zidek, T., and Schober, E. Partial r2-values based on deviance residuals in poisson regression models. *N MEDZN UND BOlOGE*, 341.

[22] Montine, T. J., Phelps, C. H., Beach, T. G., Bigio, E. H., Cairns, N. J., Dickson, D. W., Duyckaerts, C., Frosch, M. P., Masliah, E., Mirra, S. S., et al. National institute on aging–alzheimer's association guidelines for the neuropathologic assessment of alzheimer's disease: a practical approach. *Acta neuropathologica 123*, 1 (2012), 1–11.

[23] Nelson, P. T., Dickson, D. W., Trojanowski, J. Q., Jack, C. R., Boyle, P. A., Arfanakis, K., Rademakers, R., Alafuzoff, I., Attems, J., Brayne, C., et al. Limbic-predominant age-related tdp-43 encephalopathy (late): consensus working group report. *Brain 142*, 6 (2019), 1503–1527.

[24] Nichols, E., Merrick, R., Hay, S. I., Himali, D., Himali, J. J., Hunter, S., Keage, H. A. D., Latimer, C. S., Scott, M. R., Steinmetz, J. D., Walker, J. M., Wharton, S. B., Wiedner, C. D., Crane, P. K., Keene, C. D., Launer, L. J., Matthews, F. E., Schneider, J., Seshadri, S., White, L., Brayne, C., and Vos, T. The prevalence, correlation, and co-occurrence of neuropathology in old age: Harmonisation of 12 measures across six community-based autopsy studies of dementia. *The Lancet Healthy Longevity 4*, 3 (2023), e115–e125.

[25] Ovsepian, S. V., and O'Leary, V. B. Can arginase inhibitors be the answer to therapeutic challenges in Alzheimer's disease? *Neurotherapeutics 15*, 4 (2018), 1032–1035.

[26] Perez, R. K., Gordon, M. G., Subramaniam, M., Kim, M. C., Hartoularos, G. C., Targ, S., Sun, Y., Ogorodnikov, A., Bueno, R., Lu, A., et al. Single-cell rna-seq reveals cell type–specific molecular and genetic associations to lupus. *Science 376*, 6589 (2022), eabf1970.

[27] Prater, K. E., Green, K. J., Mamde, S., Sun, W., Cochoit, A., Smith, C. L., Chiou, K. L., Heath, L., Rose, S. E., Wiley, J., et al. Human microglia show unique transcriptional changes in alzheimer's disease. *Nature Aging 3*, 7 (2023), 894–907.

[28] Robinson, J. L., Xie, S. X., Baer, D. R., Suh, E., Van Deerlin, V. M., Loh, N. J., Irwin, D. J., McMillan, C. T., Wolk, D. A., Chen-Plotkin, A., et al. Pathological combinations in neurodegenerative disease are heterogeneous and disease-associated. *Brain 146*, 6 (2023), 2557–2569.

[29] Santiago, J. A., and Potashkin, J. A. The impact of disease comorbidities in Alzheimer's disease. *Frontiers in Aging Neuroscience 13* (2021), 631770.

[30] Santpere, G., Garcia-Esparcia, P., Andres-Benito, P., Lorente-Galdos, B., Navarro, A., and Ferrer, I. Transcriptional network analysis in frontal cortex in l ewy body diseases with focus on dementia with l ewy bodies. *Brain Pathology 28*, 3 (2018), 315–333.

[31] Scheltens, P., Blennow, K., Breteler, M. M., De Strooper, B., Frisoni, G. B., Salloway, S., and Van der Flier, W. M. Alzheimer's disease. *The Lancet 388*, 10043 (2016), 505–517.

[32] Siew, J. J., Chen, H.-M., Chiu, F.-L., Lee, C.-W., Chang, Y.-M., Chen, H.-L., Nguyen, T. N. A., Liao, H.-T., Liu, M., Hagar, H.-T., et al. Galectin-3 aggravates microglial activation and tau transmission in tauopathy. *The Journal of clinical investigation 134*, 2 (2024).

[33] Tan, Y., Zheng, Y., Xu, D., Sun, Z., Yang, H., and Yin, Q. Galectin-3: a key player in microglia-mediated neuroinflammation and alzheimer's disease. *Cell & bioscience 11*, 1 (2021), 78.

[34] Veitch, V., and Zaveri, A. Sense and sensitivity analysis: Simple post-hoc analysis of bias due to unobserved confounding. *Advances in neural information processing systems 33* (2020), 10999–11009.

[35] Wang, E., Yu, K., Cao, J., Wang, M., Katsel, P., Song, W.-m., Wang, Z., Li, Y., Wang, X., Wang, Q., et al. Multiscale proteomic modeling reveals protein networks driving alzheimer's disease pathogenesis. *Cell 188*, 22 (2025), 6186–6204.

[36] Wang, J., Zhao, Q., Hastie, T., and Owen, A. B. Confounder adjustment in multiple hypothesis testing. *Annals of statistics 45*, 5 (2017), 1863.

[37] Whalley, H. C., Papmeyer, M., Romaniuk, L., Johnstone, E. C., Hall, J., Lawrie, S. M., Sussmann, J. E., and McIntosh, A. M. Effect of variation in diacylglycerol kinase eta (DGKH)

gene on brain function in a cohort at familial risk of bipolar disorder. *Neuropsychopharmacology 37*, 4 (2012), 919–928.

[38] Yazar, S., Alquicira-Hernandez, J., Wing, K., Senabouth, A., Gordon, M. G., Andersen, S., Lu, Q., Rowson, A., Taylor, T. R., Clarke, L., et al. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science 376*, 6589 (2022), eabf3041.