We thank the reviewers for their feedback. We're glad the reviewers see our work as tackling an important and challenging problem and found our theory interesting and well-motivated by a real example. We appreciate that our presentation is abstract and technical, which made it difficult for R1 to follow, but as R3 pointed out is despite best efforts, the abstract nature of the presentation is potentially unavoidable given the material.

[R2: Strength of assumptions for theorem 1 and theorem 2] Both theorems rely on causal redundancy. This is an untestable assumption. However, causal redundancy is plausible in settings like GWAS where the population structure and the effects are orthogonal. See PCA correction, Price et al. and appendix B.1 of our paper for a discussion.

In theorem 1, assumption 1 has three parts: 1) that the gradient flow converges, 2) that the confounder value of the surrogate matches the confounder value whose effect is of interest, and 3) that the surrogate intervention lies in the support of the pre-outcome variables. Further, all parts can be validated on observed data. We have expanded the discussion about assumption 1. Assumption 2 is a standard technical condition required for expectations and their gradients to exist and be finite.

Coming to theorem 2, assumption 1 requires a consistent estimator of $\mathbb{E}[\mathbf{y} \mid \mathbf{t}]$, which can be provided with regression. Assumption 2 simply defines how the surrogate estimate $\hat{t}$ is obtained. Assumption 3 is a set of standard regularity assumptions which help control how the surrogate estimation error propagates to the effect error, like bounded Lipschitz-constant on the outcome and confounder functions and bounded spectral norm on the Hessian. We have added this discussion to the paper.

[R2: Comparison with prior work] Thank you for this question. To our knowledge, there exists no prior work that identifies and estimates causal effects when given a general differentiable functional confounder.

[R1: Why is positivity violated?] First note that we use $\mathbf{z} = h(\mathbf{t})$ to denote that the confounder's value $\mathbf{z}$ is provided by a "part" of the pre-outcome variables $\mathbf{t}$. Positivity is when for any $t \in \text{supp}(\mathbf{t})$, the conditional support of $\mathbf{z}$ given $\mathbf{t} = t$ exactly matches the marginal support of $\mathbf{z}$. Positivity is violated in EFC because

$$\forall t, t_2 \in \text{supp}(\mathbf{t}) \;\; s.t. \;\; h(t_2) \neq h(t) \implies p(\mathbf{z} = h(t_2) \mid \mathbf{t} = t) = 0 \neq p(\mathbf{z} = h(t_2)) > 0$$

In words, two different confounder values cannot occur for the same t. We have added this to the paper.

[R1, R2, R3: Examples in the presentation] One example of causal redundancy is the assumption underlying existing GWAS methods like the PCA correction (Price et al.). For the PCA correction, as we discuss in the appendix B.1, the functional confounder is the genotype projected onto axes of genetic variation: $h(\mathbf{t}) = A\mathbf{t}$. Then, causal redundancy is satisfied when the outcome function is $f(\mathbf{t}, h(\mathbf{t})) = A_p\mathbf{t} + h(\mathbf{t})$ and $A_p \perp A$. We have added this running example in the paper.

[R3 : Examples of surrogate interventions] Surrogate interventions are elements in the space that $\mathbf{t}$ lies in and not a distinct type of intervention. Their effects are defined like any other $t \in \text{supp}(\mathbf{t})$. For example, with $\mathbf{t}^i$ as the ith coordinate of $\mathbf{t}$ consider the SCM (as in section 2.1)

$$\mathbf{t} \in \mathbf{R}^2, \quad h(\mathbf{t}) = \mathbf{t}^1 - \mathbf{t}^2, \quad \mathbf{y} = \mathbf{t}^1 + \mathbf{t}^2 + h(\mathbf{t}).$$

Consider we want the true conditional effect for $t = [1, 1]$ and $h(t_2) = 1$:

$$\phi(t = [1, 1], h(t_2) = 1) = t^1 + t^2 + h(t_2) = 3.$$

The surrogate for this $t, h(t_2)$ would be $t' = [1.5, 0.5]$ because its confounder value and conditional effect match $h(t_2)$ and $\phi(t = [1, 1], h(t_2) = 1)$ respectively:

$$h(t') = 1.5 - 0.5 = 1 = h(t_2) \quad \text{and} \quad \phi(t', h(t')) = 3 = \phi(t = [1, 1], h(t_2) = 1).$$

[R3, Applicability of LODE] We thank the reviewer for raising an important concern about applicability.

Our work provides a formal basis to reason about the kinds of effects a practitioner can estimate given a general confounder function. Functional confounders are present in areas other than genetics. For example, consider estimating the effect of intervening on blood pressure or cholesterol on an outcome like coronary heart disease. Here a confounder would be Metabolic Syndrome which is defined as a function of cholesterol, blood pressure, blood sugar, and related measurements. Making this precise with real data would take care and is beyond the scope of the paper. Similar examples exist for other syndromes that are defined as a collection of variables that can be intervened on individually.

[R3, LODE adjusts for confounding factors] We agree with the reviewer that we do not show "correct" adjustment for confounding: this would require estimated effects as ground truth. Such GWAS ground truth effects are unavailable.

However, the SNPs we report in table 1 in the main paper and table 2 in the appendix have all been reported as relevant to celiac disease in established GWAS studies. The SNP rs2237236 has a 0 Lasso coefficient meaning that its effect would be 0 without correction, going against its established relevance. With LODE's correction, we recover this SNP as relevant. We believe that such correction atop our recovery of relevant SNPs is evidence that LODE corrects for confounding to an extent sufficient to separate relevant SNPs from irrelevant ones. We will change the language to make to make this more clear.