We thank the reviewers for their comments and valuable feedback. In the following we address their questions:

Reviewer1: We thank the Reviewer for several useful suggestions that we will gladly include in the revisions.
Q1: If it is the same MMD that is used for evaluation, then the better results for USD are not so surprising since the evaluation metric is part of the model.
A: To make comparisons fair, the MMD used for evaluation is different from the one used at training. For evaluation, we used a fixed MMD defined in lines 196-198: "In all our experiments we report the MMD distance with a Gaussian kernel, computed using random Fourier features (RF) with 300 RF and a kernel bandwith equal to $\sqrt{d}$".
Q2: The difference with unbalanced gradient flows is unclearly stated in my opinion.
A: Thank you for pointing out the lack of clarity about this. We will clarify this point in the revisions. The main difference is that unbalanced gradient flows of WFR give a full path from source to target, that requires to solve a sequential planning problem where all distributions on the path have to be simultaneously computed. Unbalanced Sobolev Descent on the other hand finds a path from source to target computing each step in a greedy way.
Q3: The proposed gradient flow algorithm shares some similarities with splitting scheme.
A: Thank you for pointing this out, this is indeed the case. We will add a discussion about this.
Q4: I would like to see the impact of the choice of the regularizing MMD.
A: Thank you for the suggestion. We will be glad to add this ablation experiment for the synthetic examples.

Reviewer2: Thank you for suggestions that we will also address in the revised paper.
Q: Explain what is the processing time of the proposed algorithms in relation to the compared methods.
A: Thank you for pointing out an opportunity to clarify our work. We gave computational complexity of Neural USD in lines 178-182, and said that it scales linearly in samples: $O(N)$. But we will be glad to give more explicit comparisons to the OT Baselines in the revisions. Entropic regularized unbalanced OT solved with Sinkhorn that were used in the paper are expensive computationally as they scale as $O(N^2)$, where N is number of samples. Better scaling alternatives in OT with sample size complexity that rivals our method ($O(N)$) exist, but either require approximation or more complex optimization schemes. Examples are Altschuler et al. that relies on a Nystrom approximation, and Genevay et al. that relies on stochastic optimization. We will add this discussion to the paper.

Reviewer3: Thank you for your feedback and useful literature pointers that we will gladly add to the revised paper.
Q1: This paper's idea is very closely related to Sobolev descent paper. The only difference is that the metric is changed from Wasserstein to Wasserstein-Fisher-Rao metric.
A: That is indeed the main formal difference compared to Sobolev descent. This slight formal change however results in several qualitative differences such as the fact that corresponding transport now include a reaction term, in addition to advection. The main practical implications are that we can therefore diffuse to and from distributions of different mass, we can deal with weighted samples, and that the descent is provably accelerated, which we also show empirically.
Q2: Is there any practical guidance of the choice of kernels in RKHS?
A: For the kernel one can use the consolidated heuristic to select the bandwidth of the Gaussians based on the median distance between samples in the source and target distributions.
Q3: The title is a bit confusing. A: Thanks for pointing this out. We'll try to highlight the Sobolev-Fisher Discrepancy.
Q4: Additional references. A: Thanks for all the suggested citations, we will make sure to add them to the main text.

Reviewer4: We thank the Reviewer for their time. We believe we can address most misunderstandings by R4.
Q1: There is no quantitative results to assess the performance of the method. A: We respectfully disagree. Figs 1b, 2b, give quantitative results, i.e. the MMD of the resulting distribution found with USD to the target. In Fig 3, the MMD to the target is also given under each image. Fig 4 quantitatively evaluates OT vs USD on a cell biology dataset.
Q2: The single-cell dataset is treated without any serious scientific considerations.
A: This is a theoretical and methodological paper. We don't think it is our place to draw scientific conclusions about biology experiments. This is the role of domain experts. Our goal was to demonstrate the potential of our method to analyse this type of data, and for that we carefully followed the same experimental protocol of Schiebinger et al.
Q3: There is not a single figure about computational time, size of the data.
A: We assume R4 might have overlooked section "Computational and Sample Complexities" (line 178) where we discuss this. We'll gladly add explicit comparisons with OT that tends to scale worse than our method (in fact, the comment on line 212 singled out by R4 was about OT, known to require subsampling for feasibility, unlike our method).
Q4: A very partial piece of code is written in the text. A: We'll release code and analyses upon acceptance.
Q5: The author speaks jargon, e.g. "witness function". A: We respectfully point out that "witness function" is an accepted term, commonly adopted in IPM papers and now permeating even the GANs literature.
Q6: this is a theoretical development but without a real substantial practical numerical/methodological input.
A: We disagree. We introduce a novel kernelized discrepancy and show how its flow can be used in practical settings for unbalanced transport. Our algorithm is numerically faster than conventional OT methods and previous alternatives.