



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Πολυτεχνική Σχολή
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τηλεπικοινωνιών

Multi-task learning in perturbation modeling

Διπλωματική Εργασία
του
Θεόδωρου Κατζάλη

Επιβλέπων: Όνομα Επίθετο
Καθηγητής Α.Π.Θ.

January 14, 2025

Abstract

abstract

Abstract

abstract

Ευχαριστίες

Τίτλος διπλωματικής

Όνομα Επίθετο
empty@auth.gr

January 14, 2025

Περιεχόμενα

1 Benchmarking	2
1.1 Datasets	3
1.1.1 Nault et al. 2022	3
1.1.2 PBMC dataset	10
1.2 Nault all cell types evaluation	13
1.2.1 Multiple doses	13
1.2.2 Single dose	14
1.2.3 Comparison	15
1.3 Nault liver cell types evaluation	29
1.3.1 Multiple doses	29
1.3.2 Single dose 30 $\mu g/kg$	30
1.3.3 Comparison	31
1.3.4 Παρατηρήσεις	45
1.4 PBMC	46
1.4.1 Comparison	47
A Ακρωνύμια και συντομογραφίες	48

Chapter 1

Benchmarking

1.1 Datasets

1.1.1 Nault et al. 2022

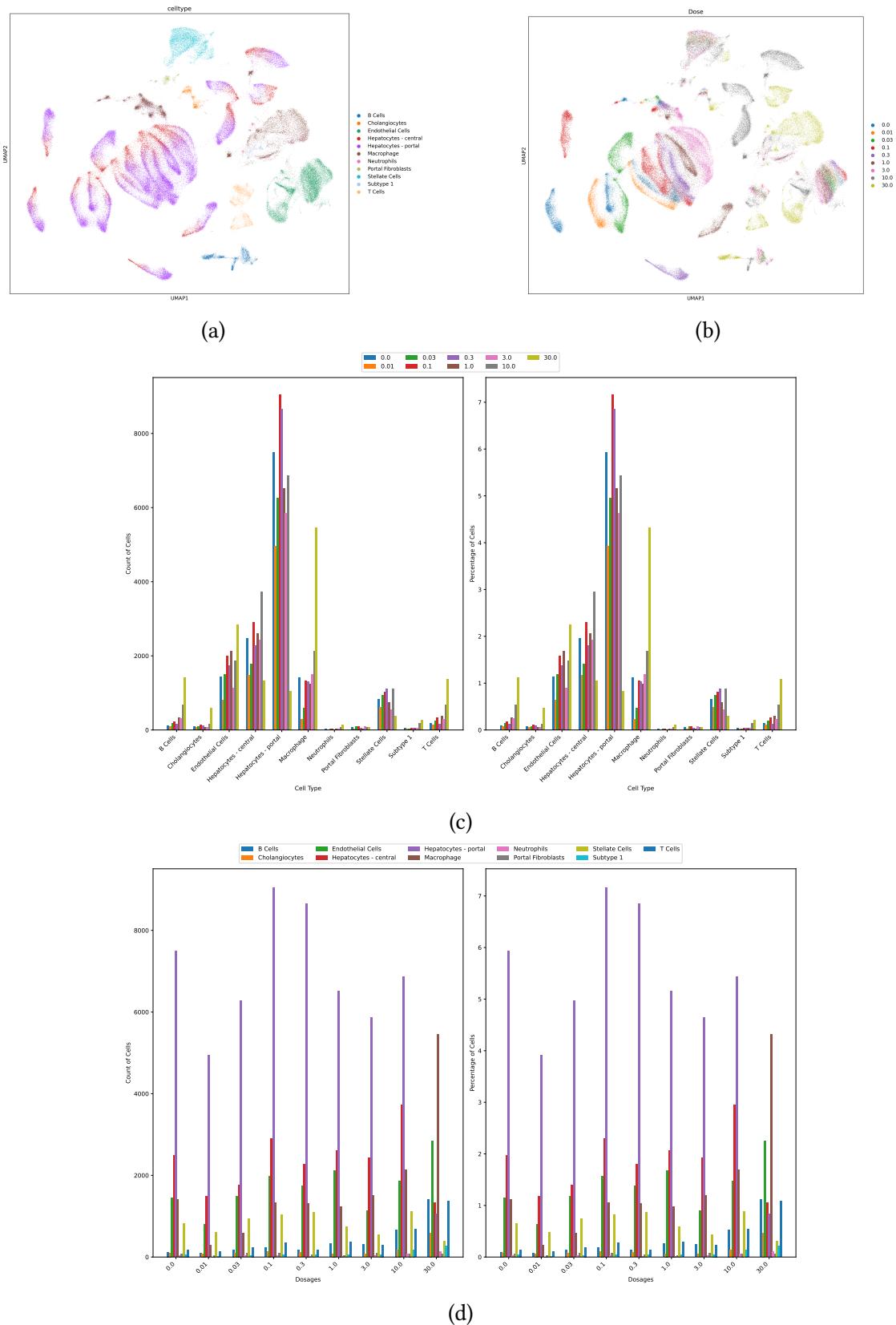


Figure 1.1: Nault overview

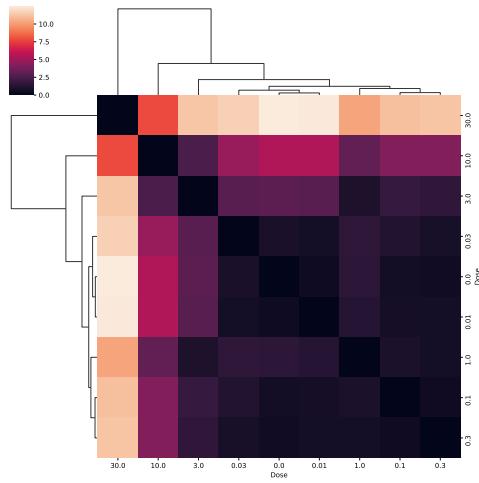


Figure 1.2: E-distance

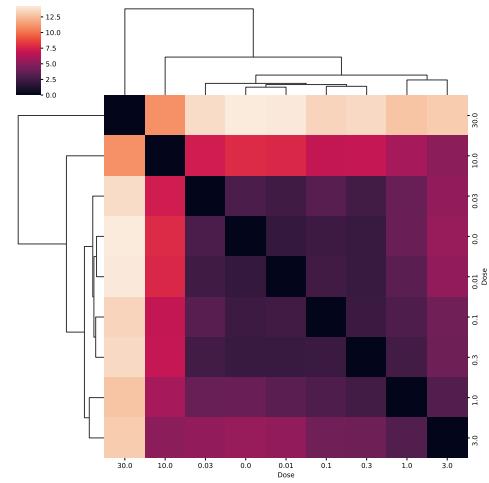


Figure 1.3: Euclidean

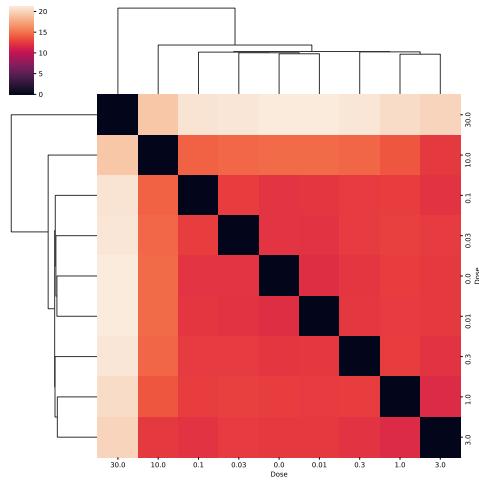


Figure 1.4: Mean pairwise

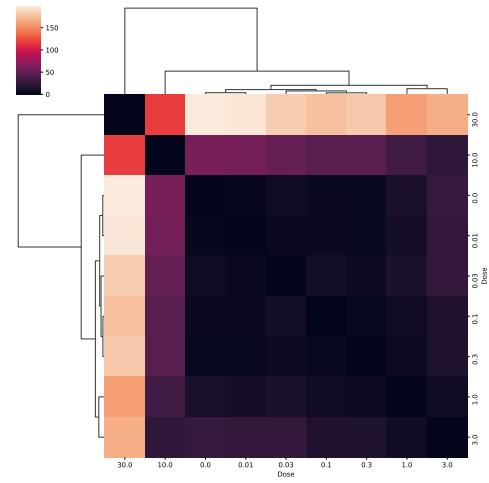


Figure 1.5: MMD

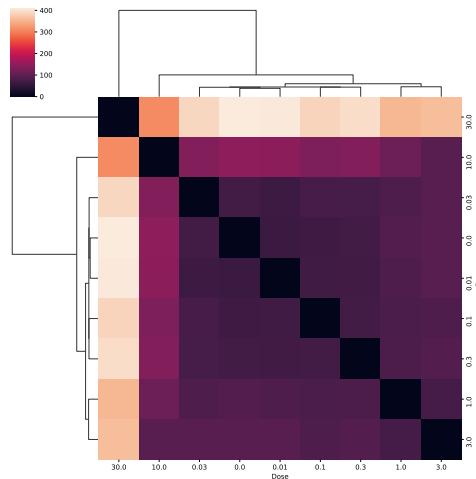


Figure 1.6: Wasserstein

Figure 1.7: Distance metrics across all cell types per dosage

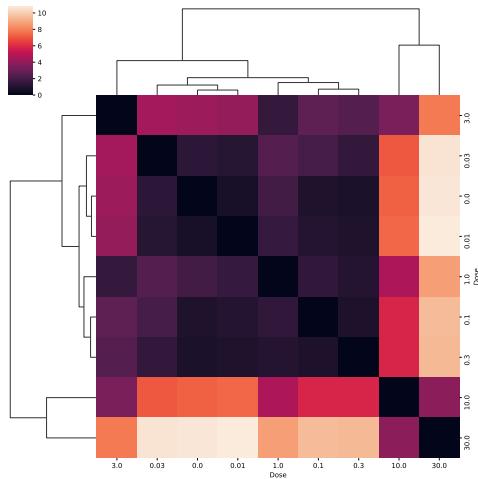


Figure 1.8: E-distance

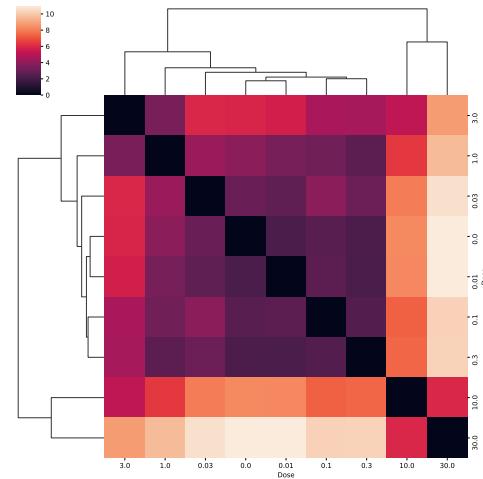


Figure 1.9: Euclidean

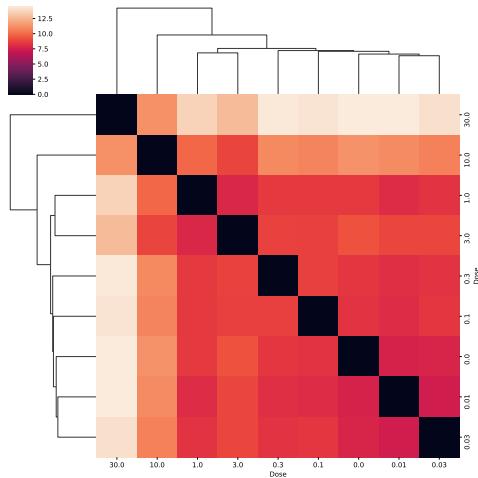


Figure 1.10: Mean pairwise

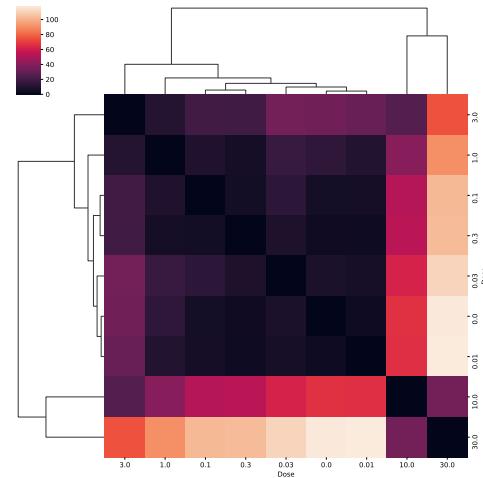


Figure 1.11: MMD

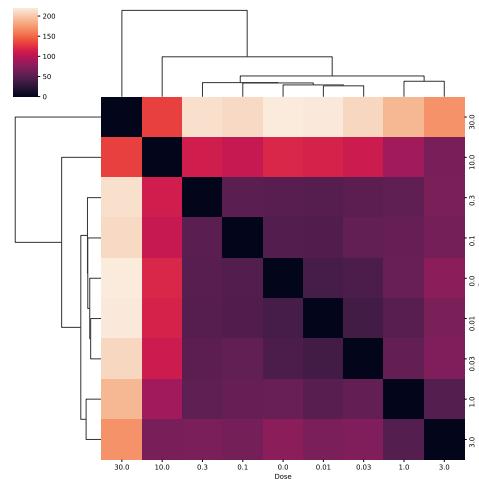


Figure 1.12: Wasserstein

Figure 1.13: Distance metrics for cell type Hepatocytes - portal per dosage

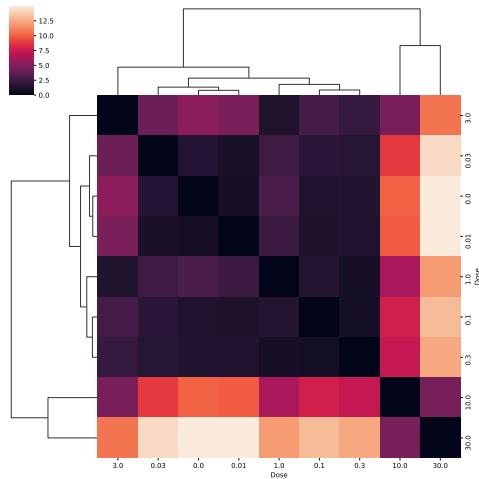


Figure 1.14: E-distance

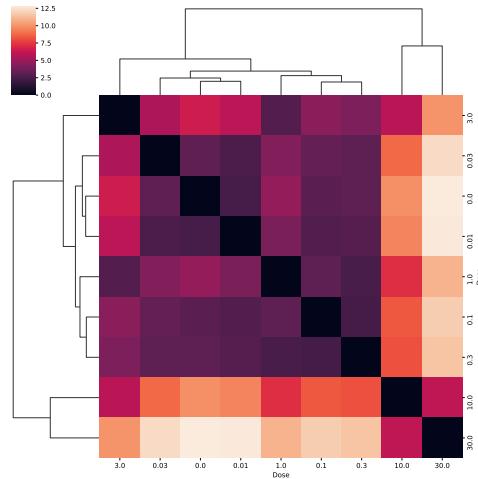


Figure 1.15: Euclidean

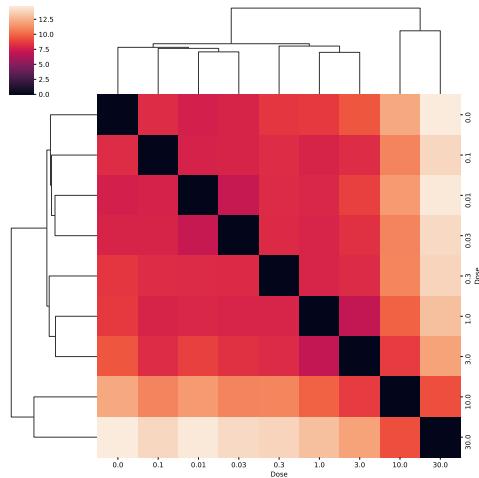


Figure 1.16: Mean pairwise

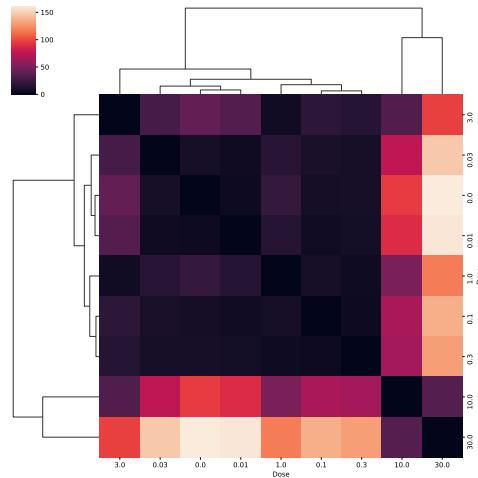


Figure 1.17: MMD

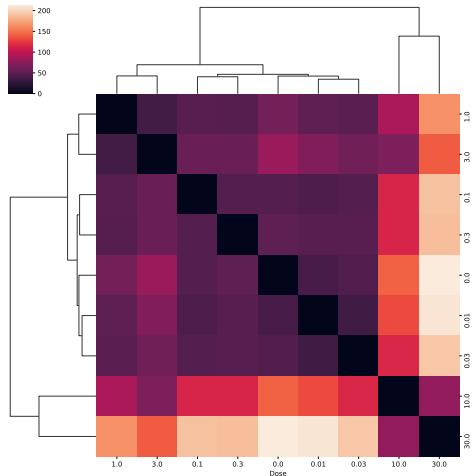


Figure 1.18: Wasserstein

Figure 1.19: Distance metrics for cell type Hepatocytes - central per dosage

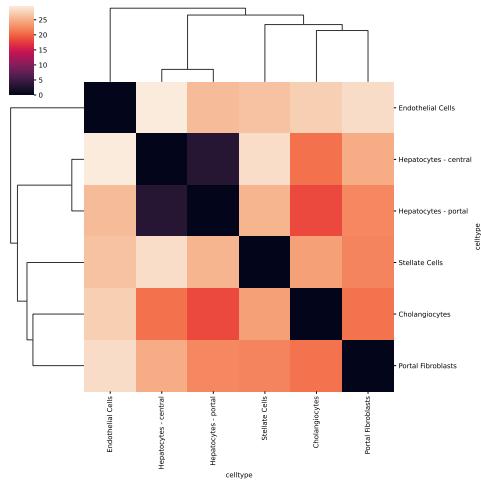


Figure 1.20: E-distance

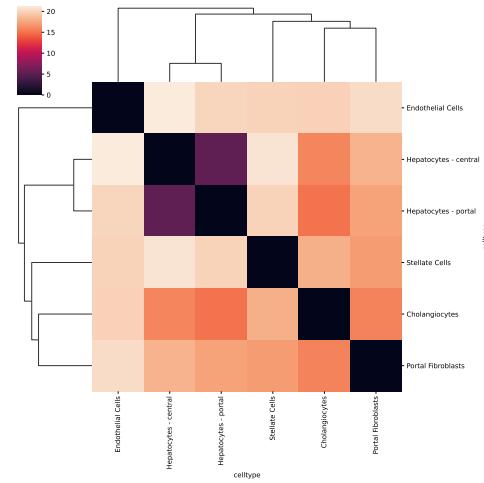


Figure 1.21: Euclidean

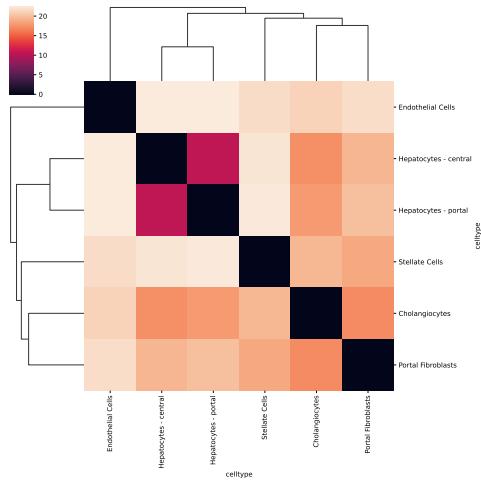


Figure 1.22: Mean pairwise

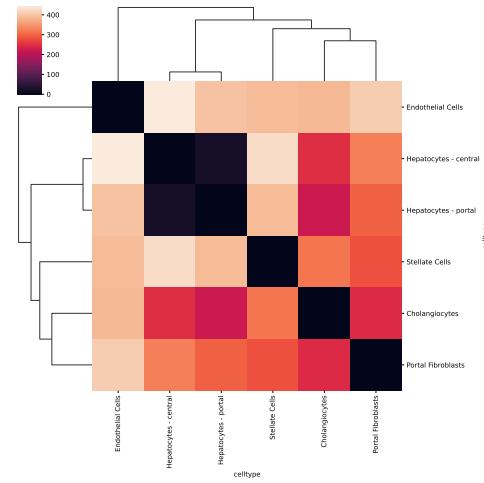


Figure 1.23: MMD

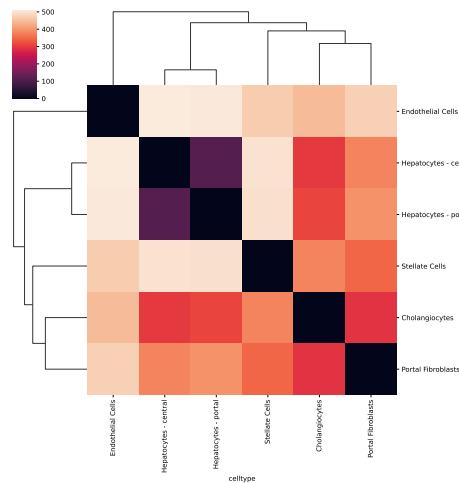


Figure 1.24: Wasserstein

Figure 1.25: Distance metrics for dosage highest $30 \mu\text{g}/\text{kg}$ per cell type

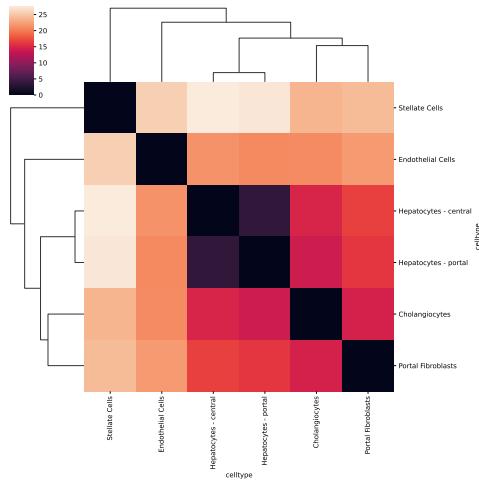


Figure 1.26: E-distance

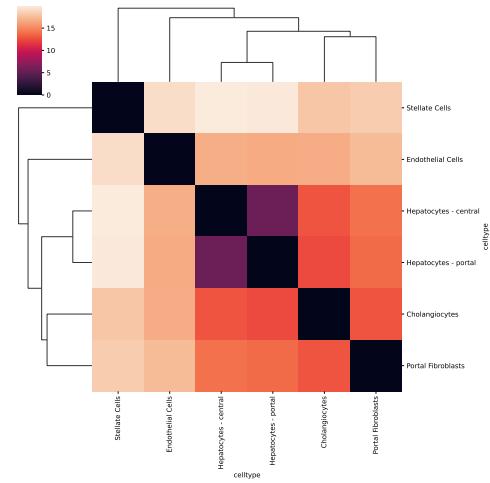


Figure 1.27: Euclidean

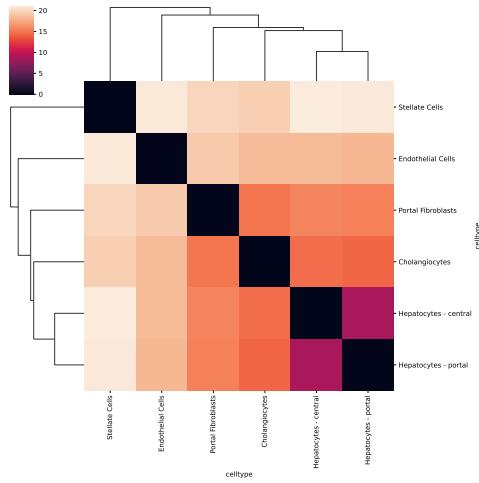


Figure 1.28: Mean pairwise

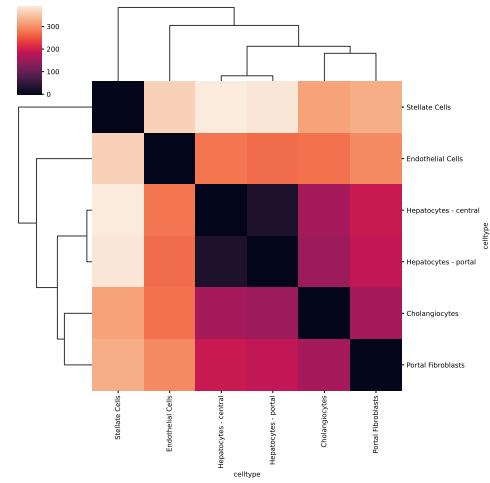


Figure 1.29: MMD

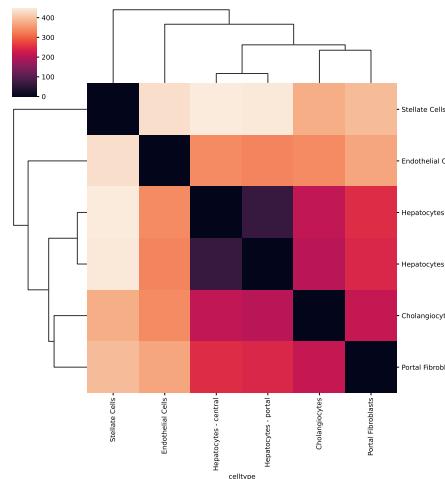


Figure 1.30: Wasserstein

Figure 1.31: Distance metrics for lowest dosage $0.01 \mu\text{g}/\text{kg}$ per cell type

1.1.2 PBMC dataset

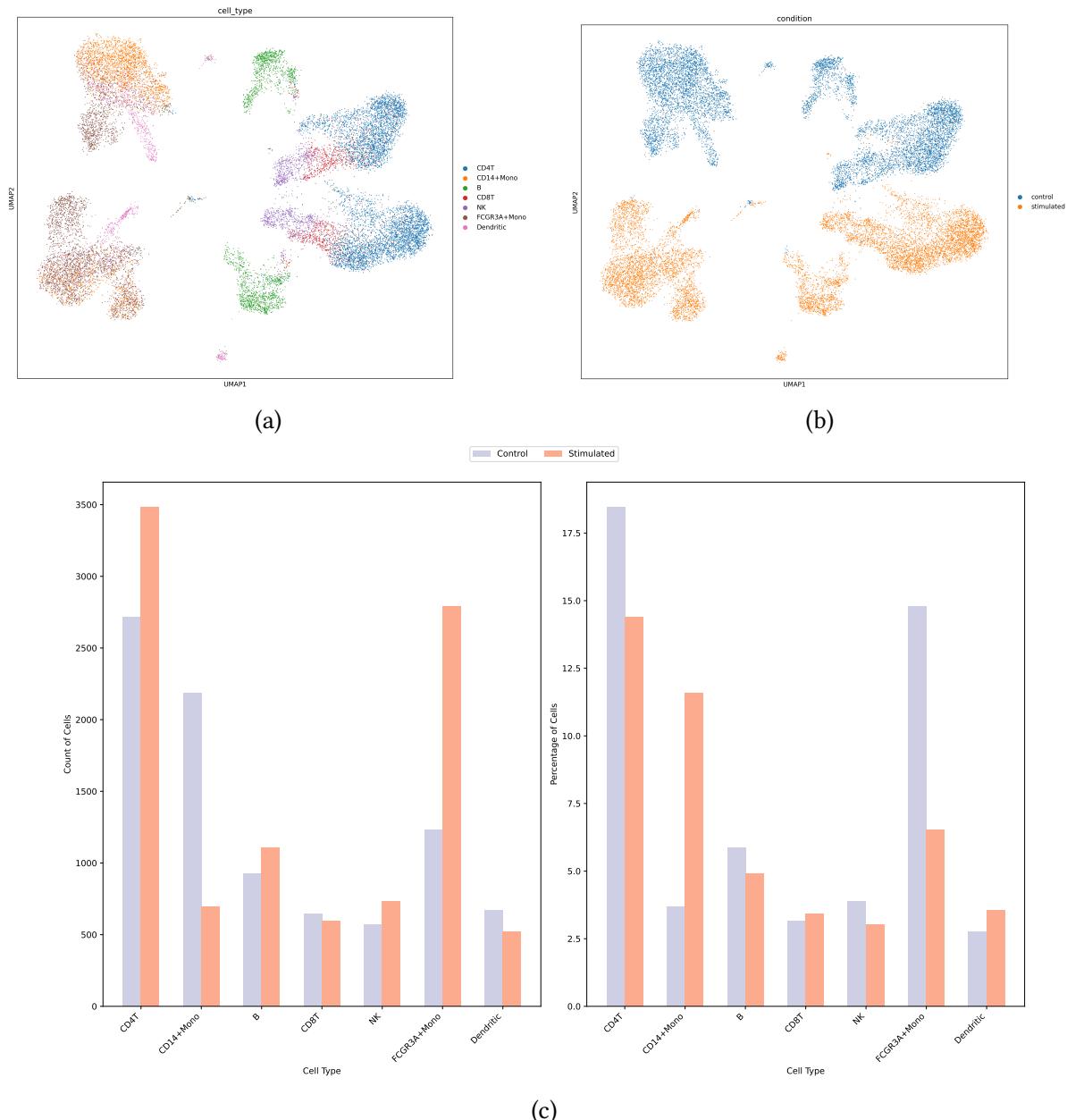


Figure 1.32: PBMC overview

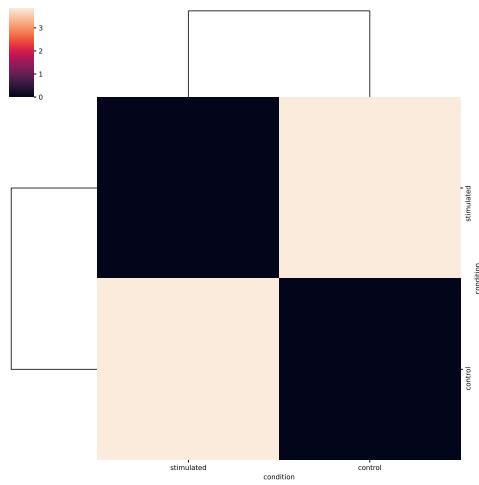


Figure 1.33: E-distance

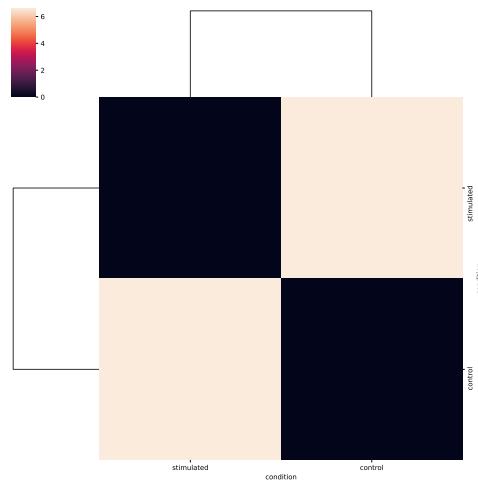


Figure 1.34: Euclidean

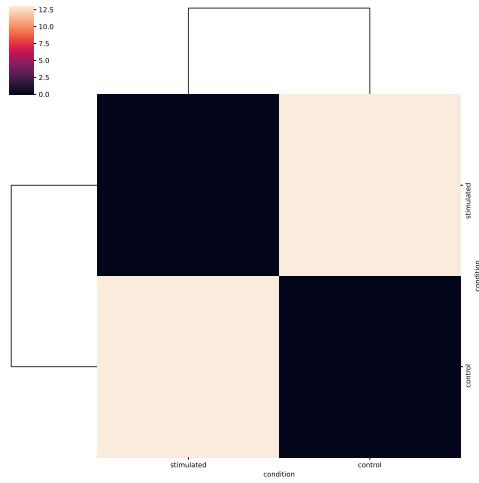


Figure 1.35: Mean pairwise

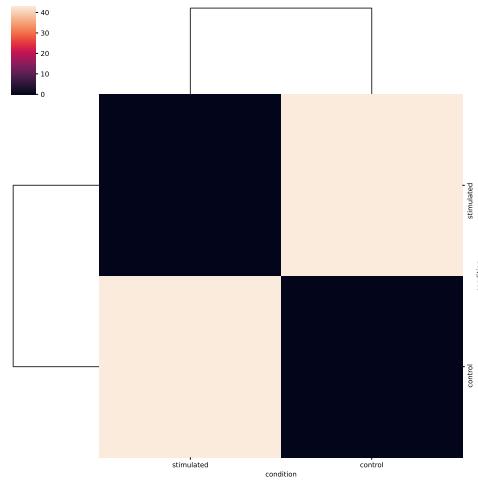


Figure 1.36: MMD

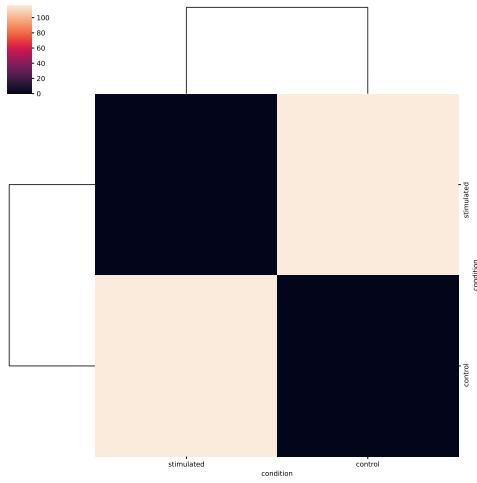


Figure 1.37: Wasserstein

Figure 1.38: Distance metrics per condition

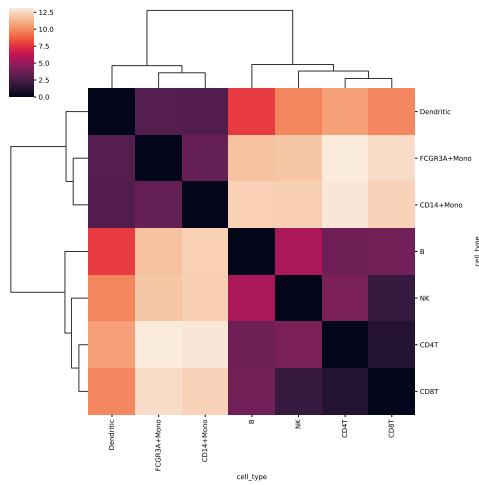


Figure 1.39: E-distance

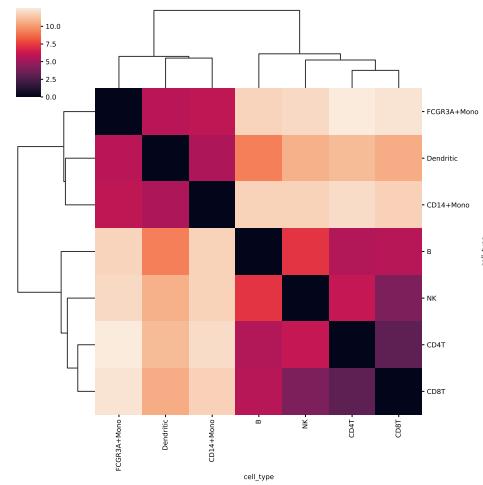


Figure 1.40: Euclidean

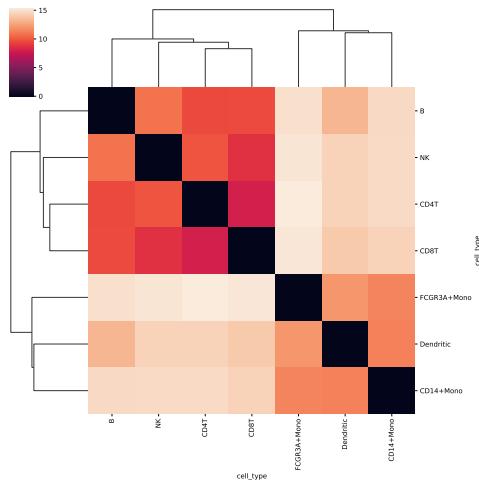


Figure 1.41: Mean pairwise

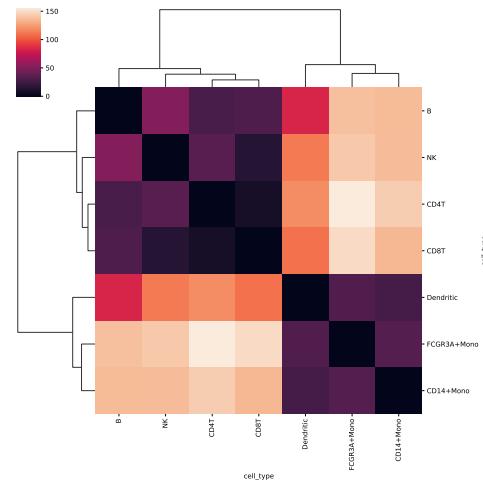


Figure 1.42: MMD

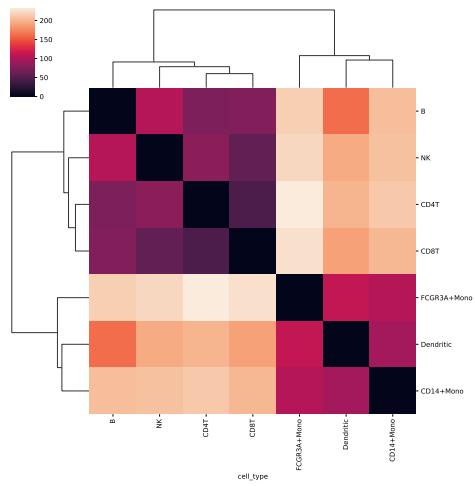
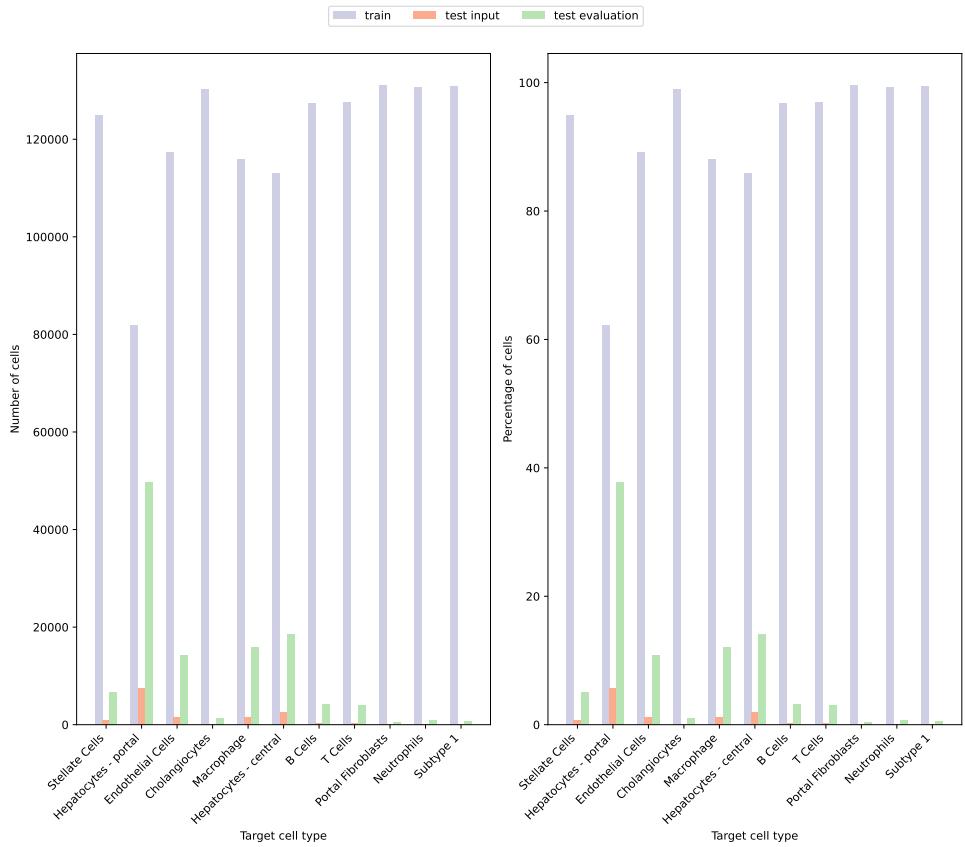
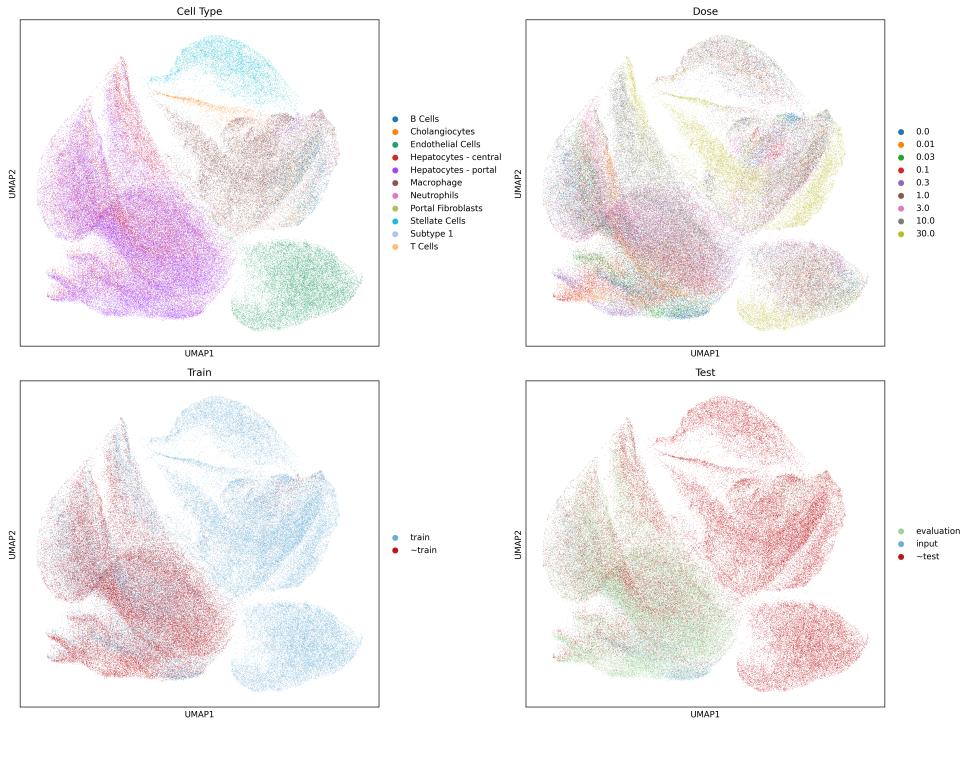


Figure 1.43: Wasserstein

Figure 1.44: Distance metrics per cell type

1.2 Nault all cell types evaluation

1.2.1 Multiple doses



1.2.2 Single dose

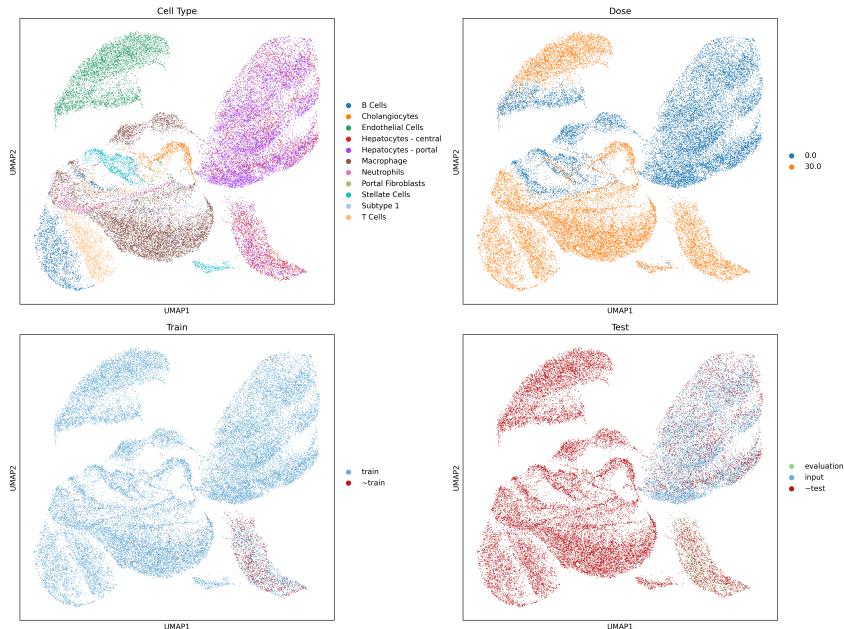


Figure 1.45: Example of $30\mu\text{g}/\text{kg}$

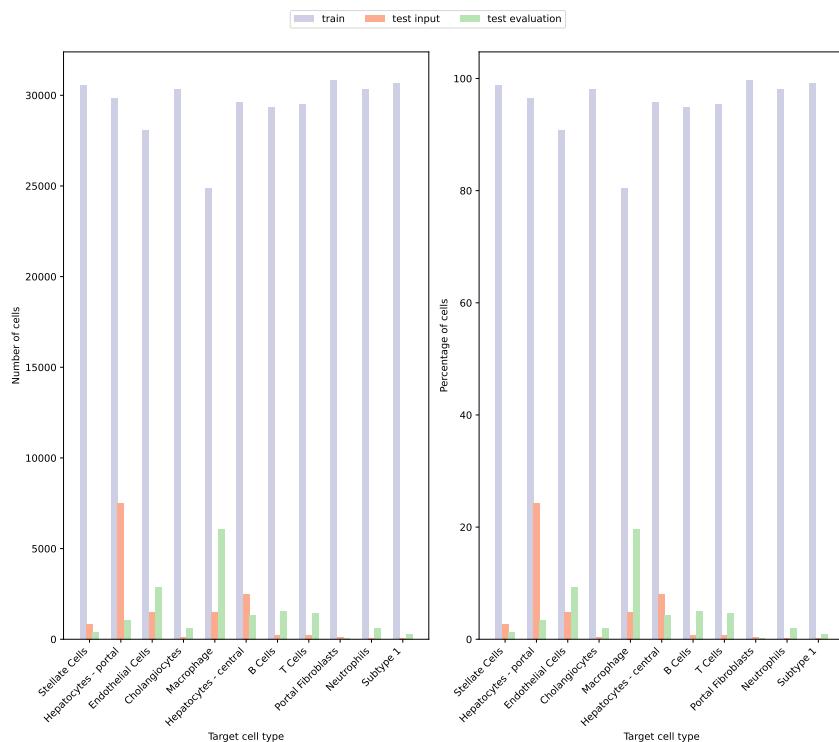


Figure 1.46: Number of cells per cell type for $30\mu\text{g}/\text{kg}$

1.2.3 Comparison

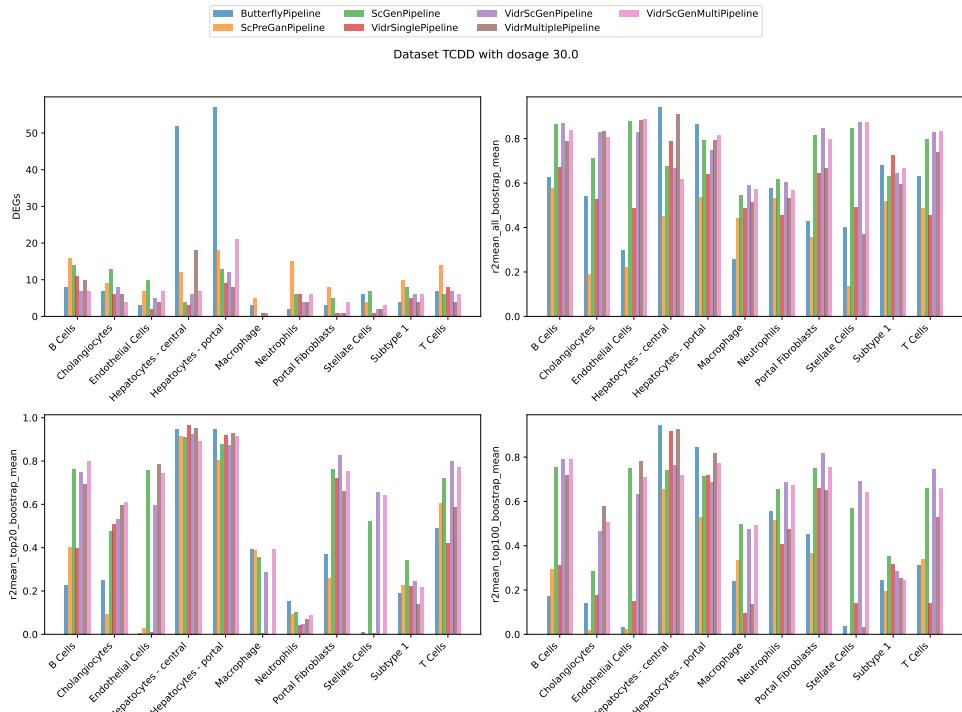


Figure 1.47: Baseline metrics for highest dosage $30 \mu\text{g}/\text{kg}$ across cell types

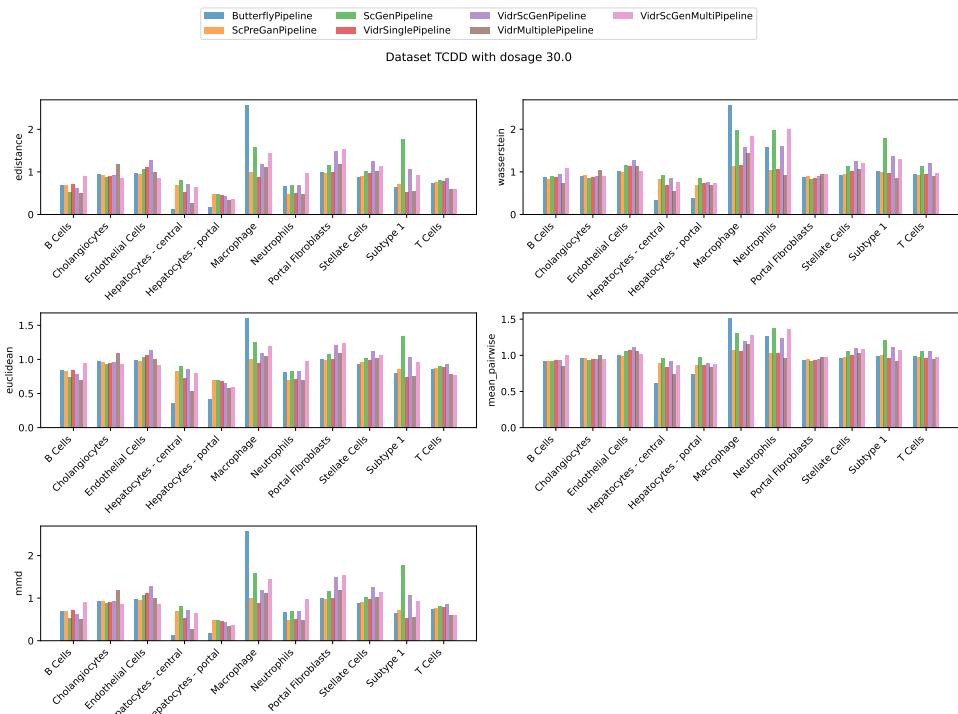


Figure 1.48: Distance metrics for highest dosage $30 \mu\text{g}/\text{kg}$ across cell types

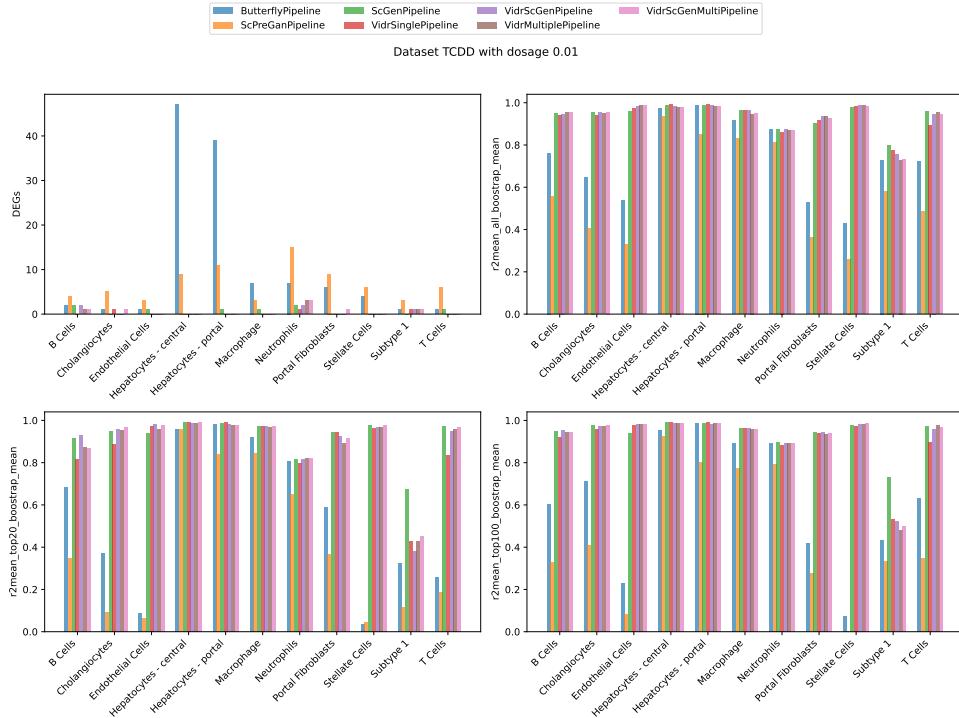


Figure 1.49: Baseline metrics for lowest dosage $0.01 \mu\text{g}/\text{kg}$ across cell types

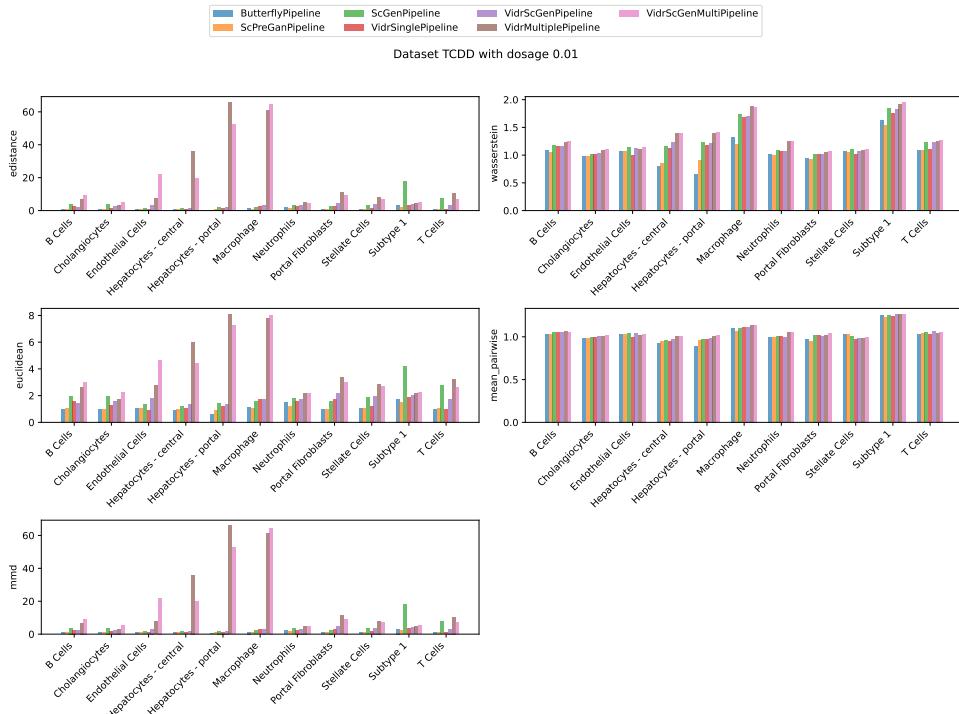


Figure 1.50: Distance metrics for lowest dosage $0.01 \mu\text{g}/\text{kg}$ across cell types

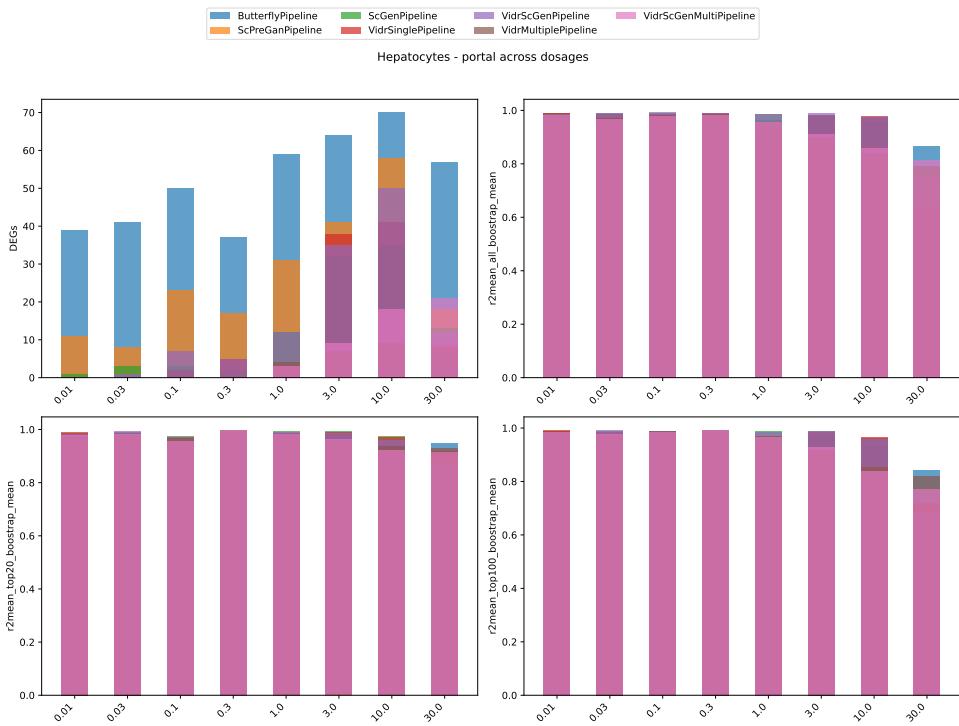


Figure 1.51: Baseline metrics for Hepatocytes - portal across dosages

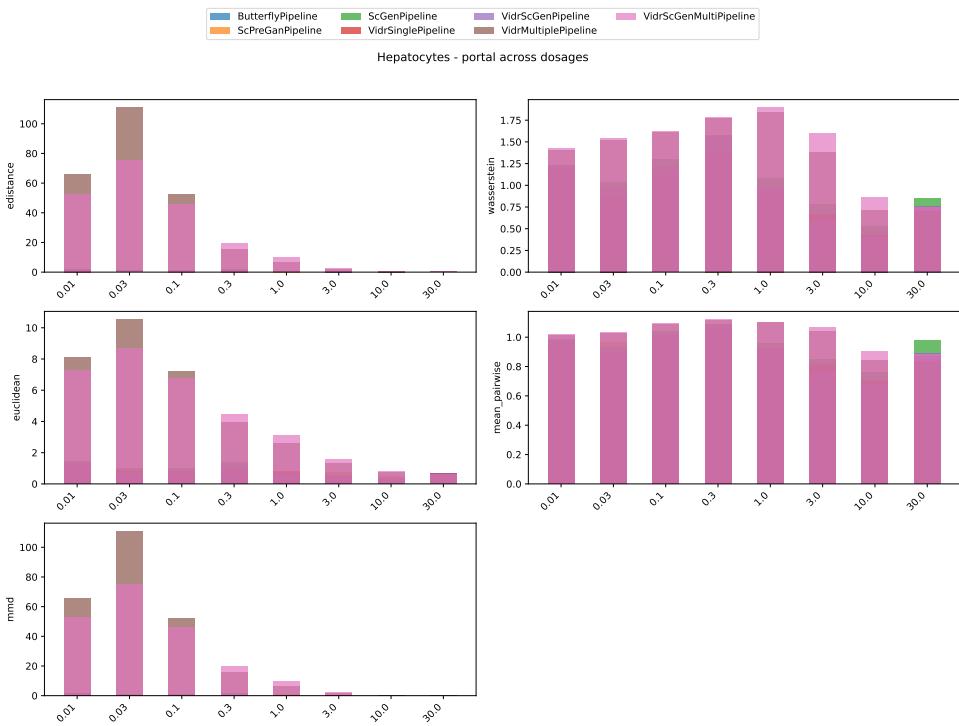


Figure 1.52: Distance metrics for Hepatocytes - portal across dosages

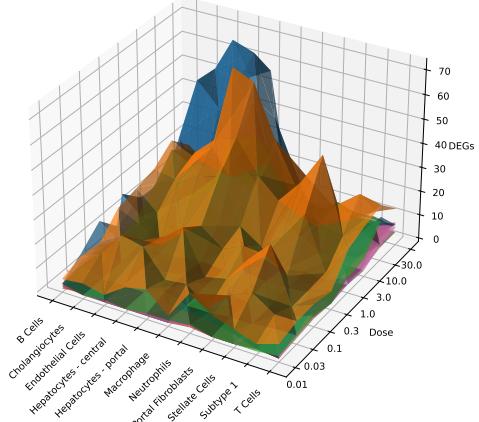


Figure 1.53

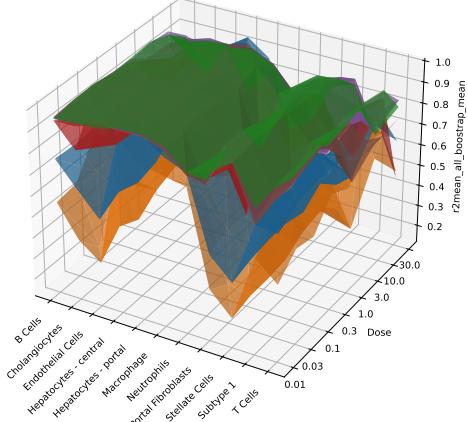


Figure 1.54

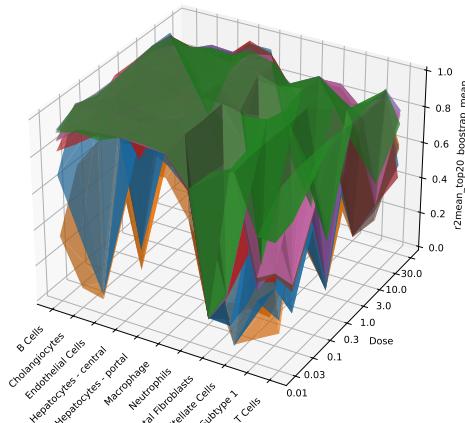


Figure 1.55

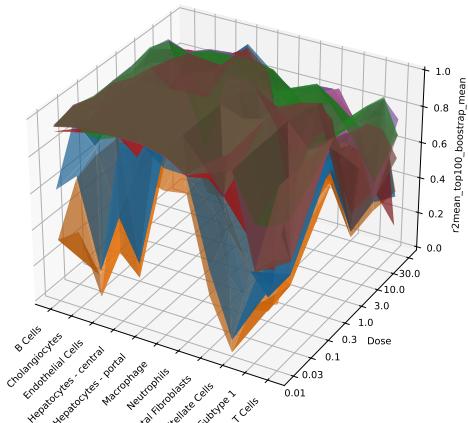


Figure 1.56

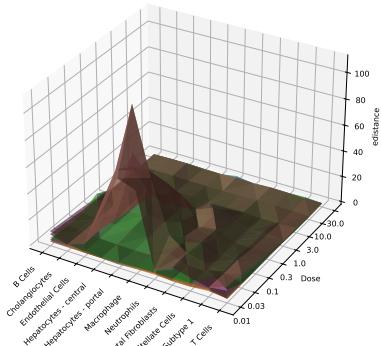


Figure 1.57: E-distance

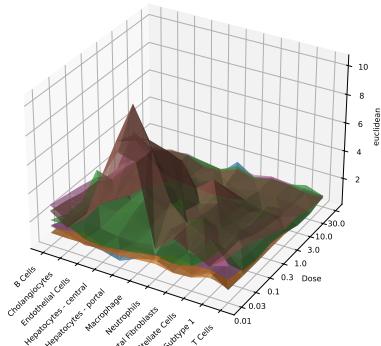


Figure 1.58: Euclidean

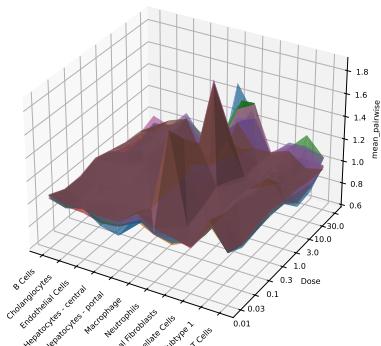


Figure 1.59: Mean pairwise

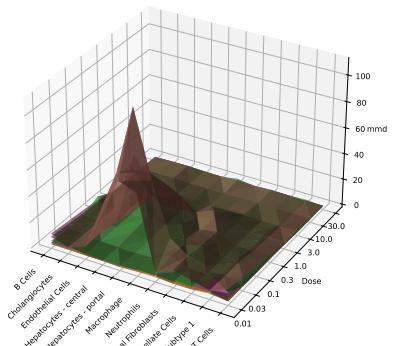
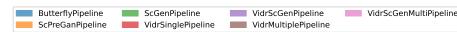


Figure 1.60: MMD

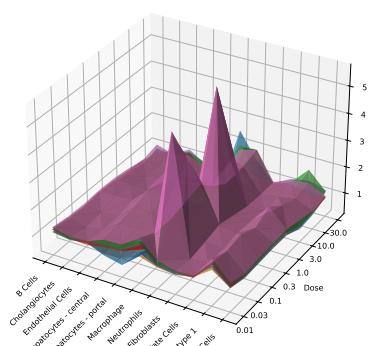
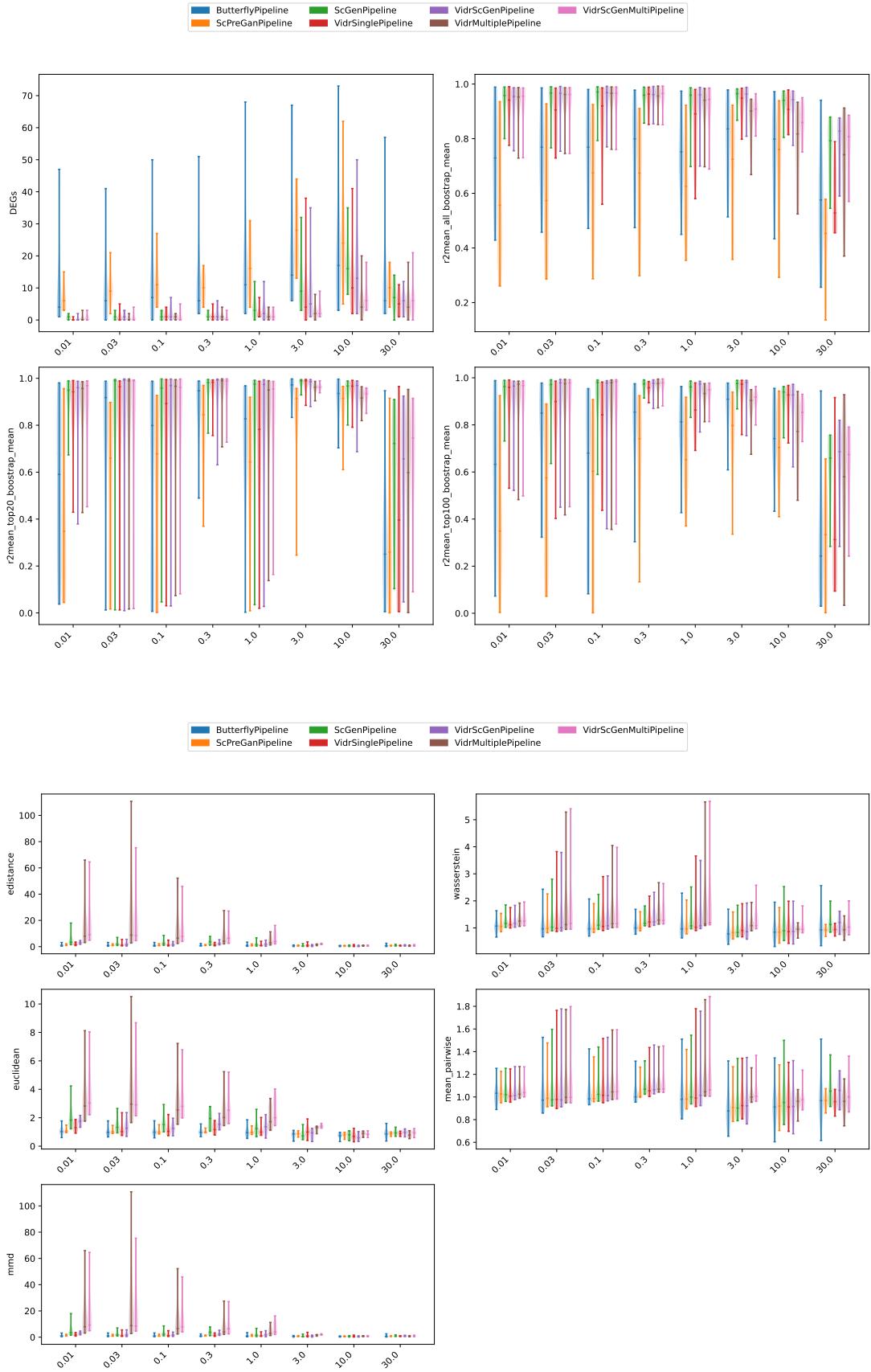
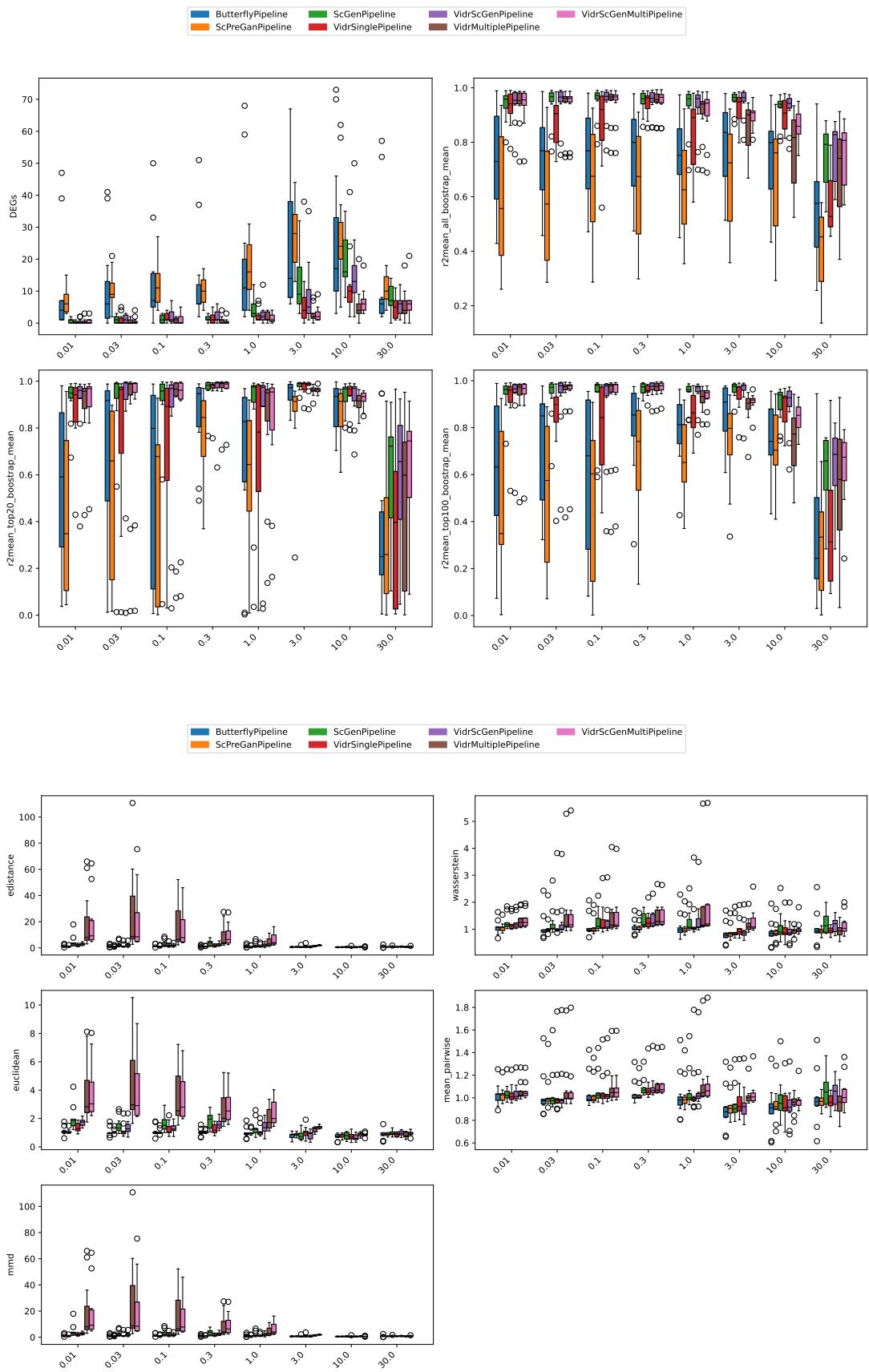
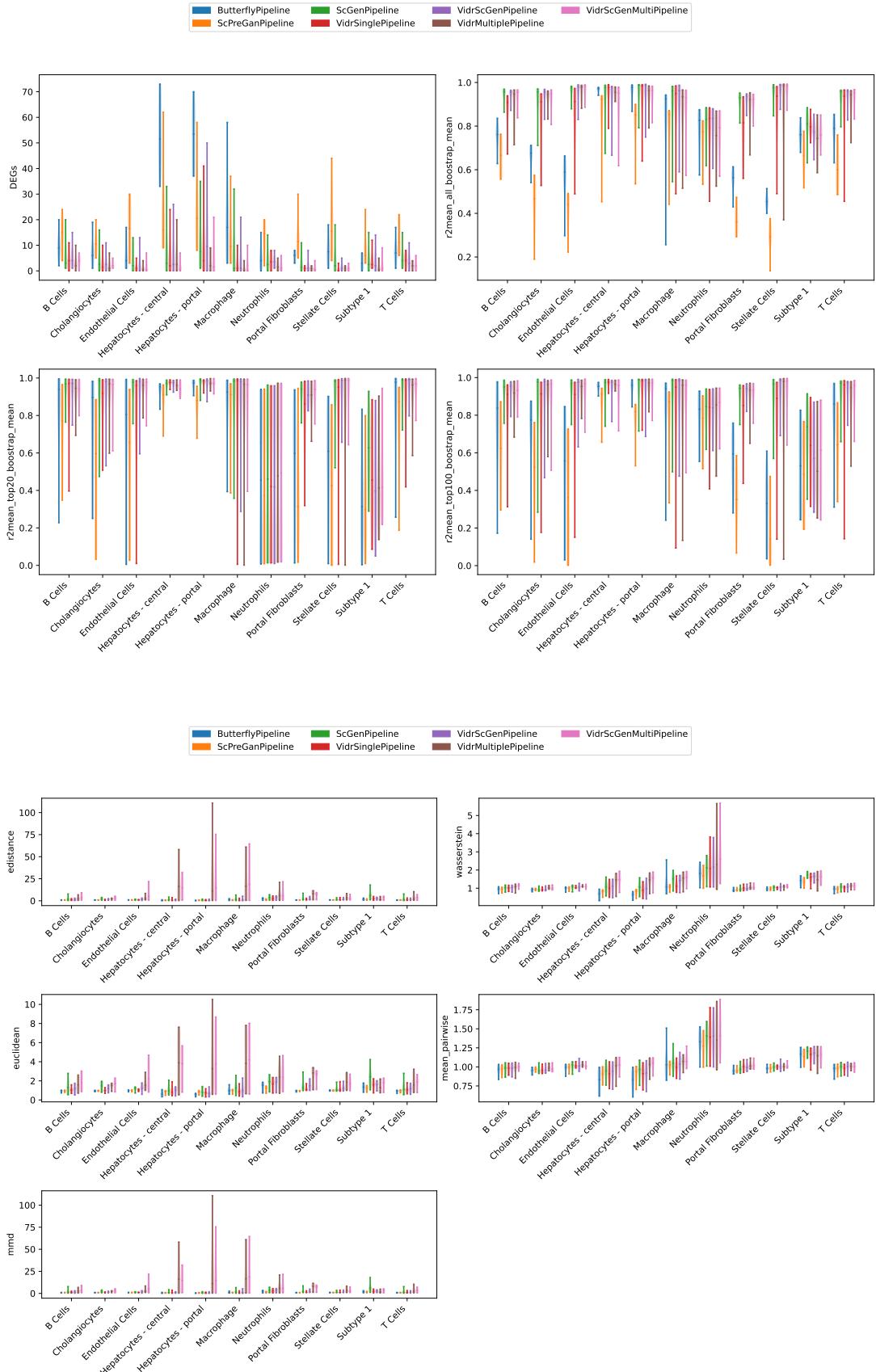


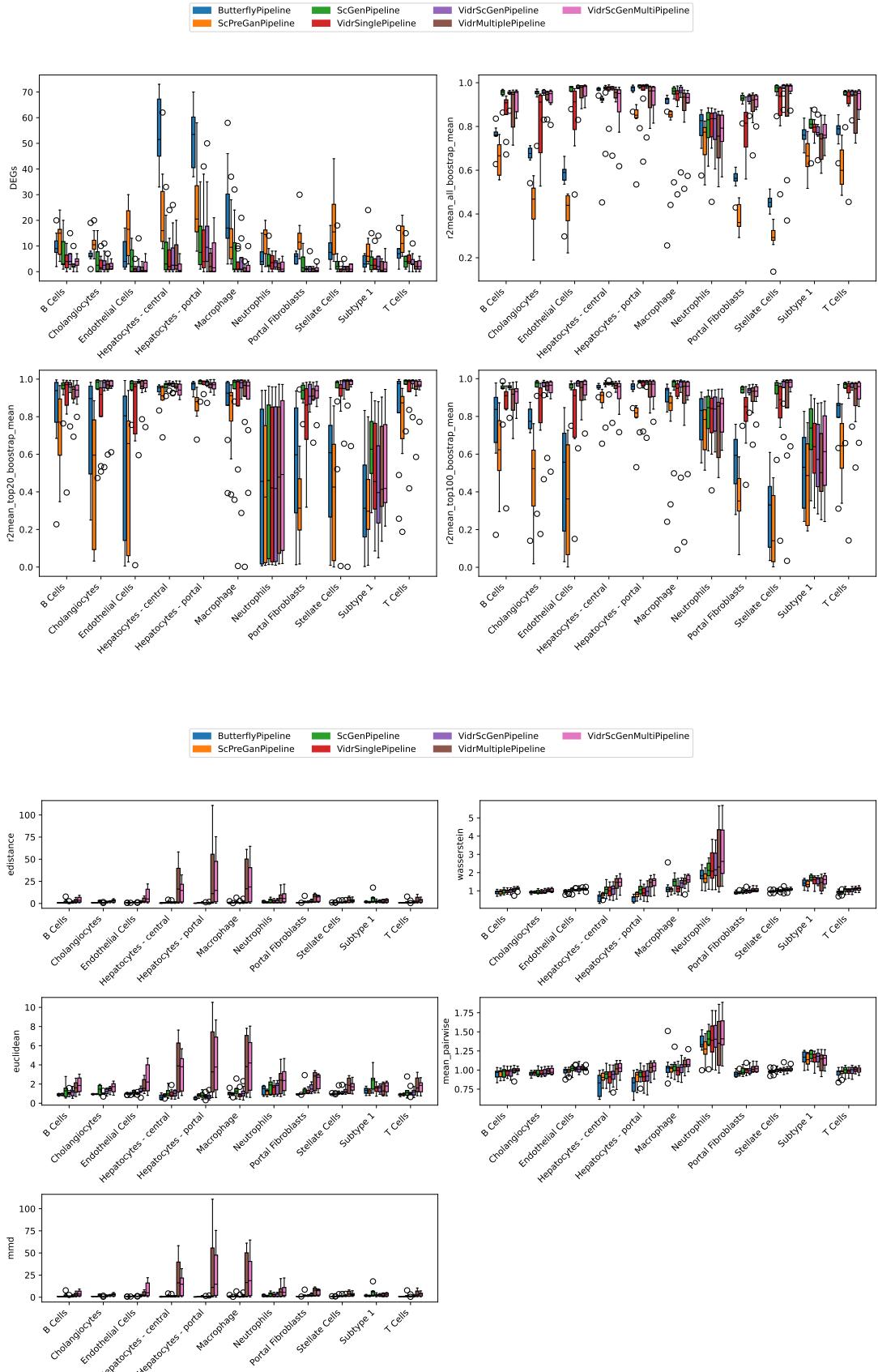
Figure 1.61: Wasserstein

Figure 1.62: Distance metrics per cell type









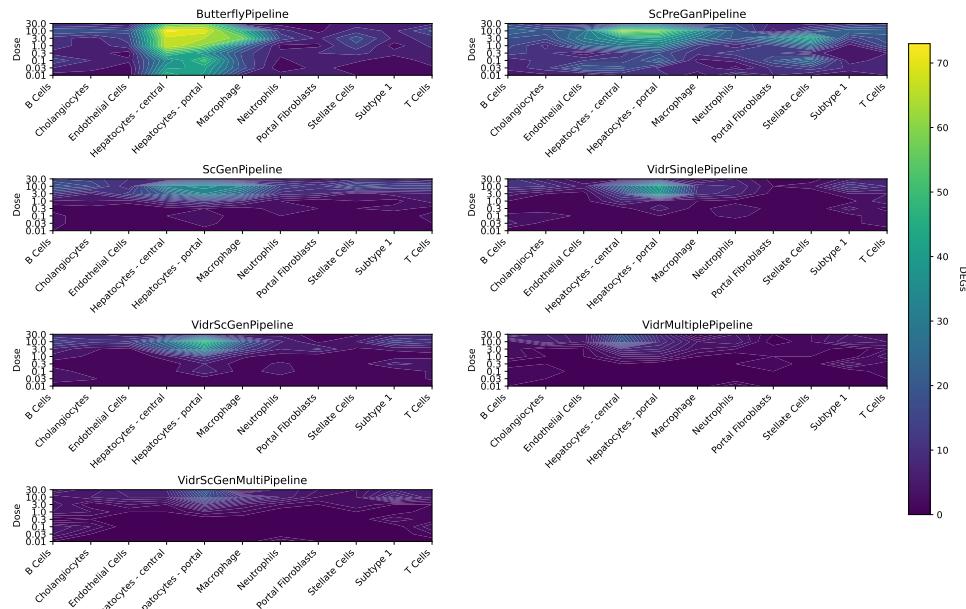


Figure 1.63: DEGs

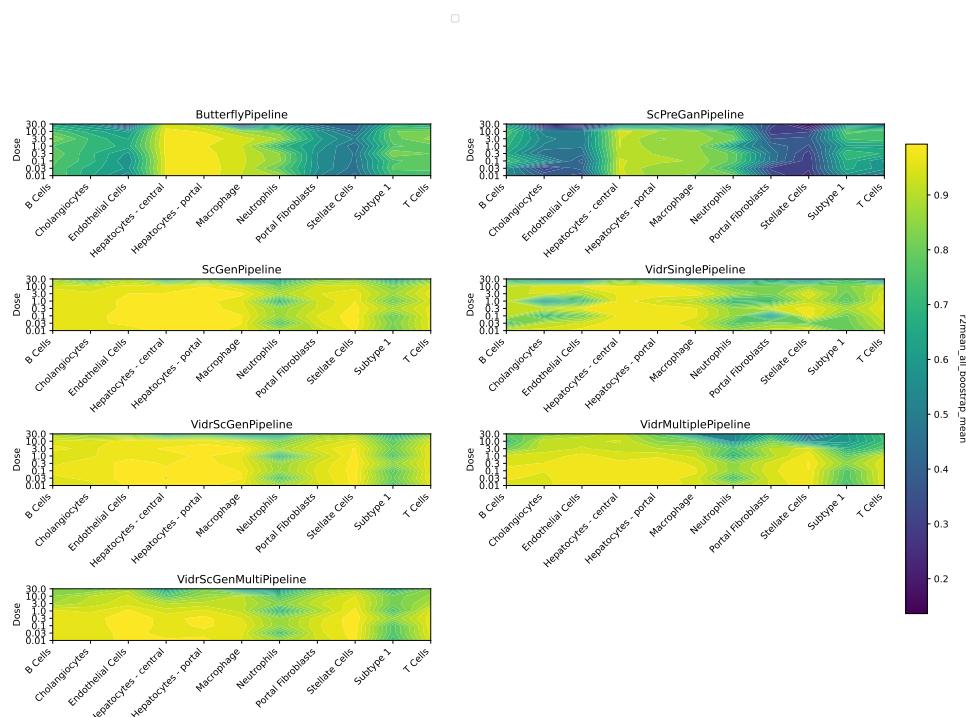


Figure 1.64: r^2 HVGs

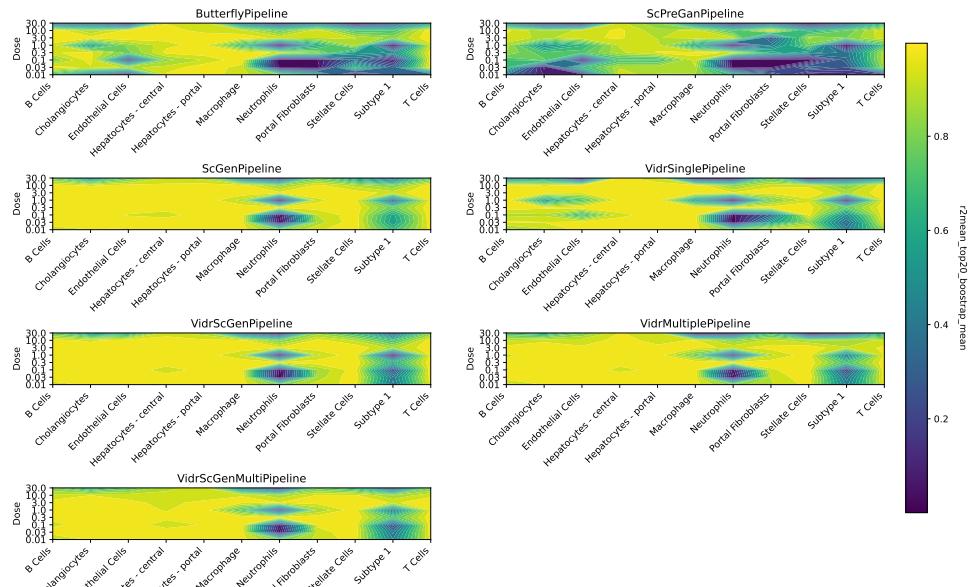


Figure 1.65: r2 top 20

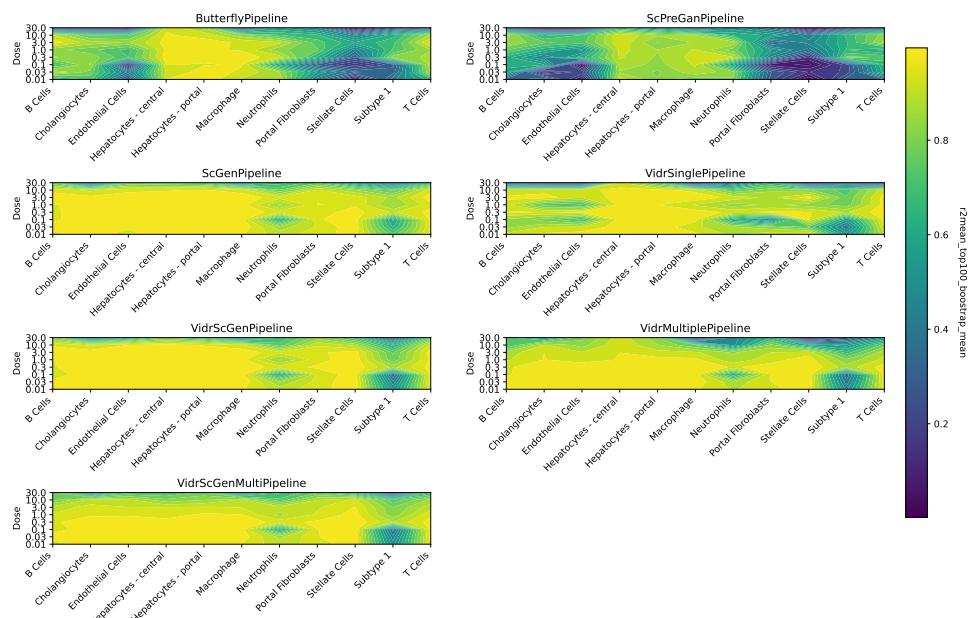


Figure 1.66: r2 top 100

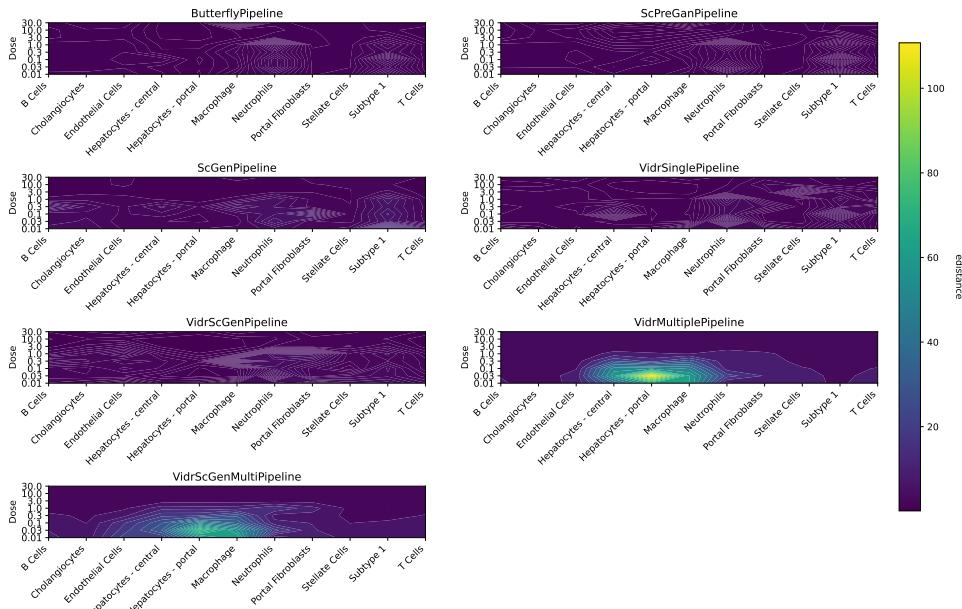


Figure 1.67: E-distance

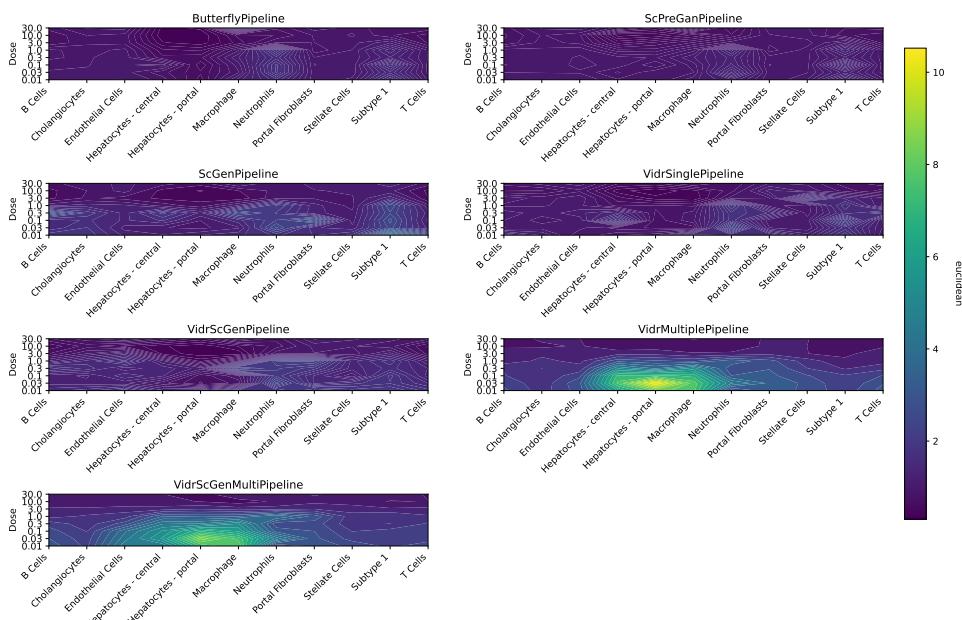


Figure 1.68: Euclidean

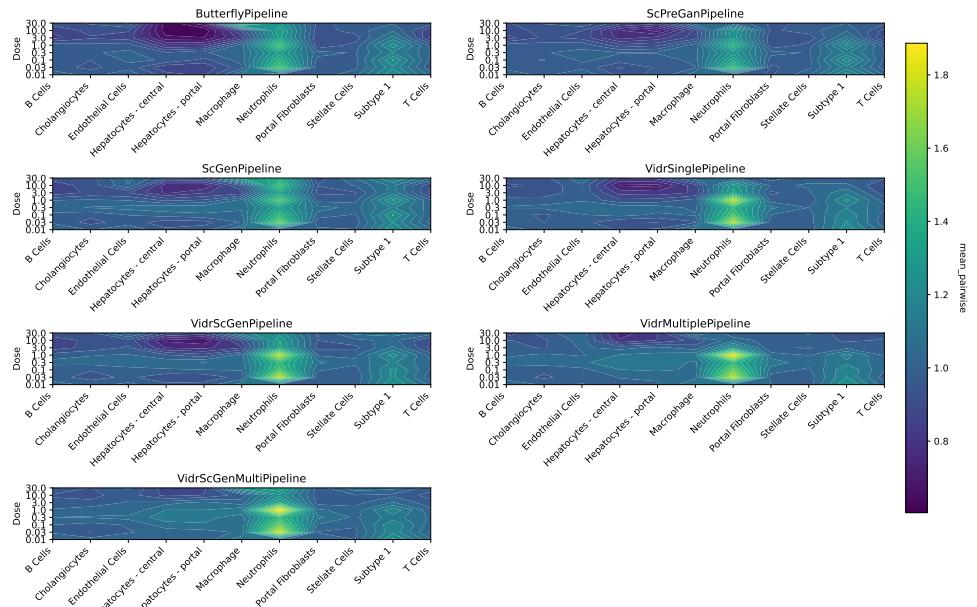


Figure 1.69: Mean pairwise

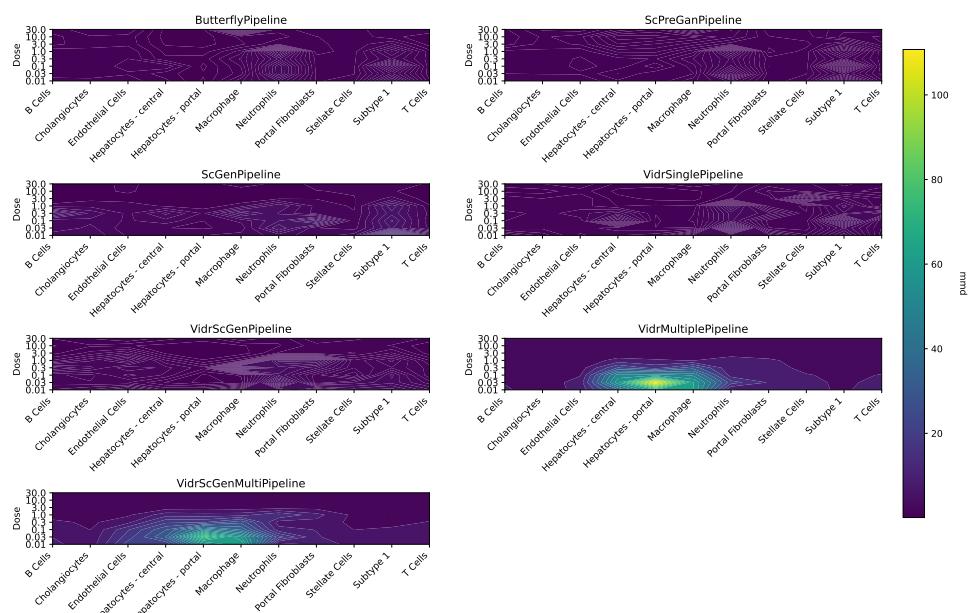


Figure 1.70: MMD

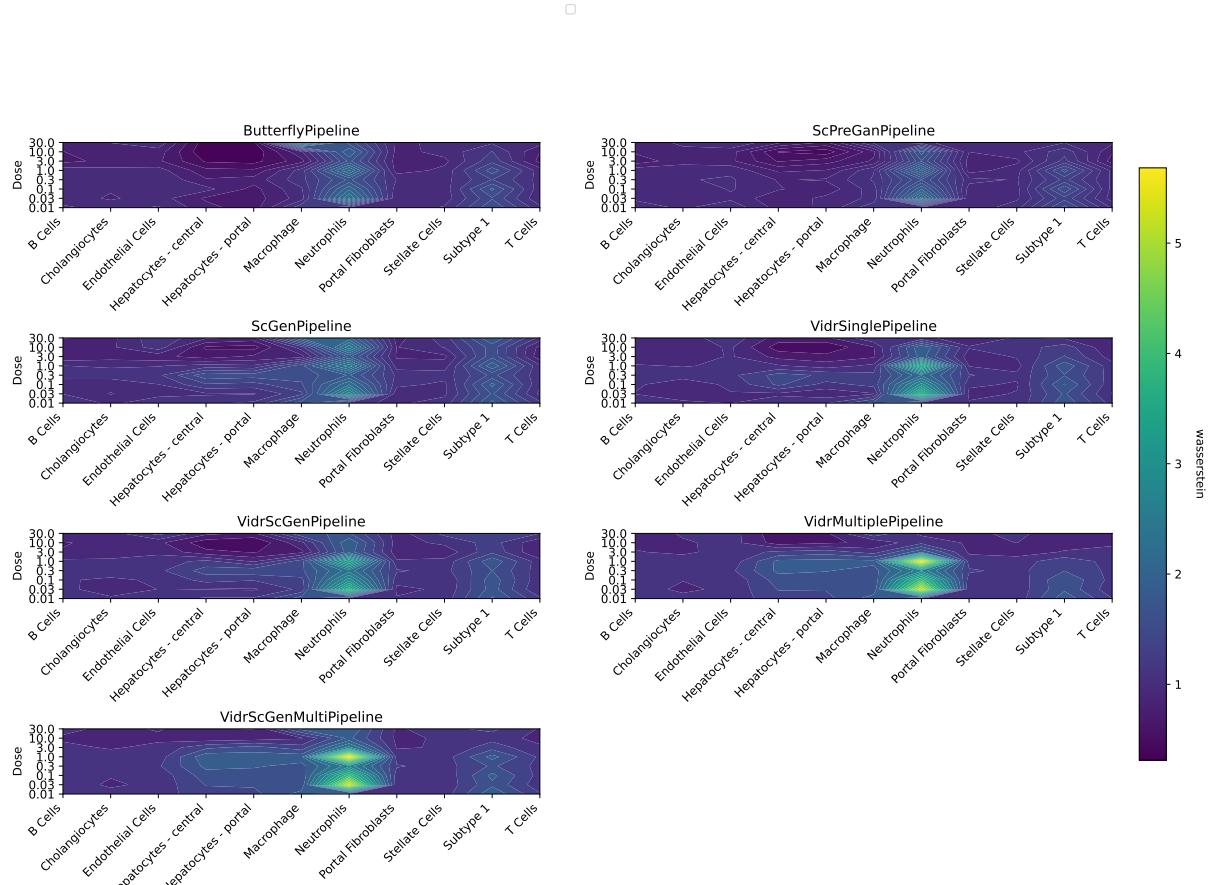
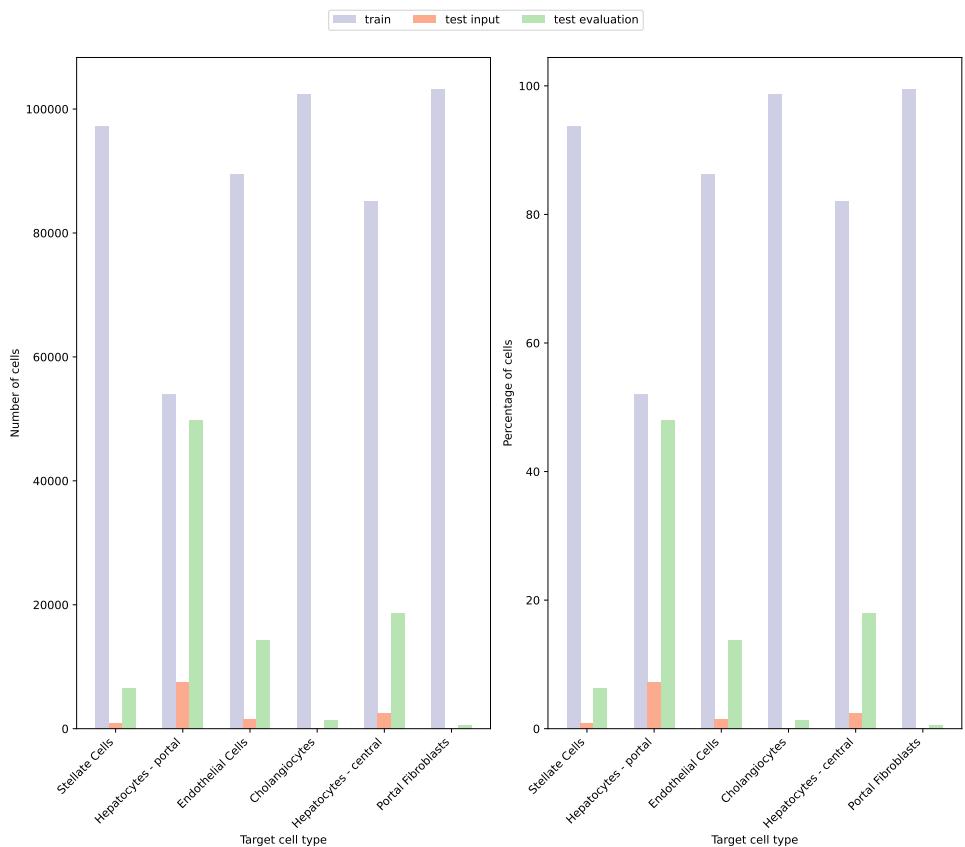
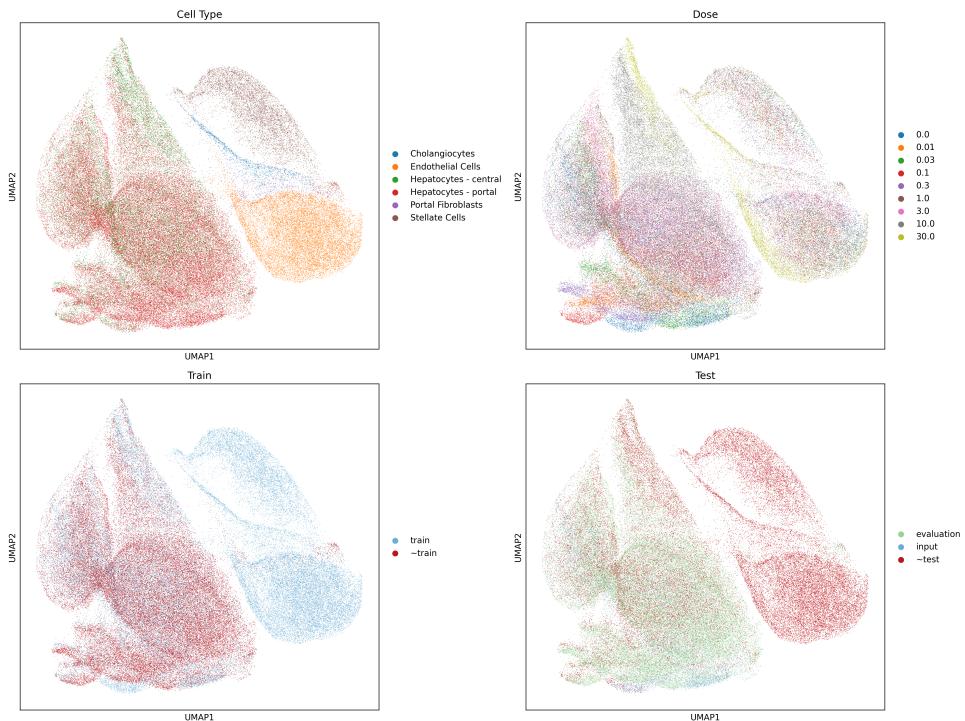


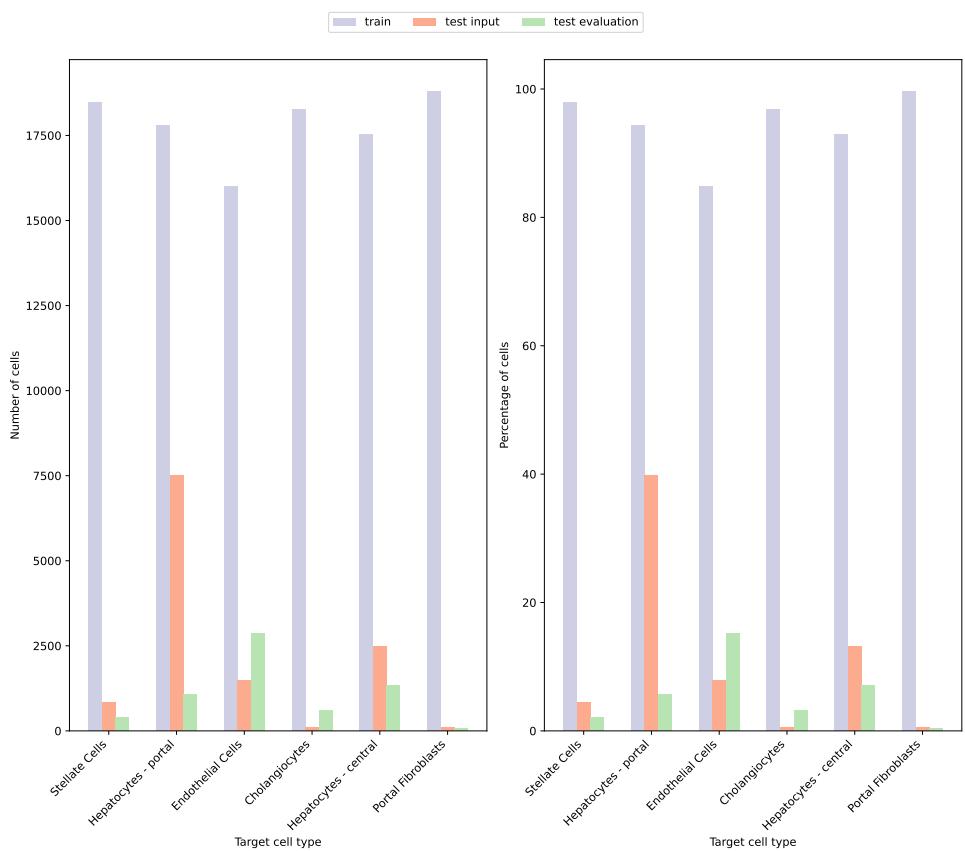
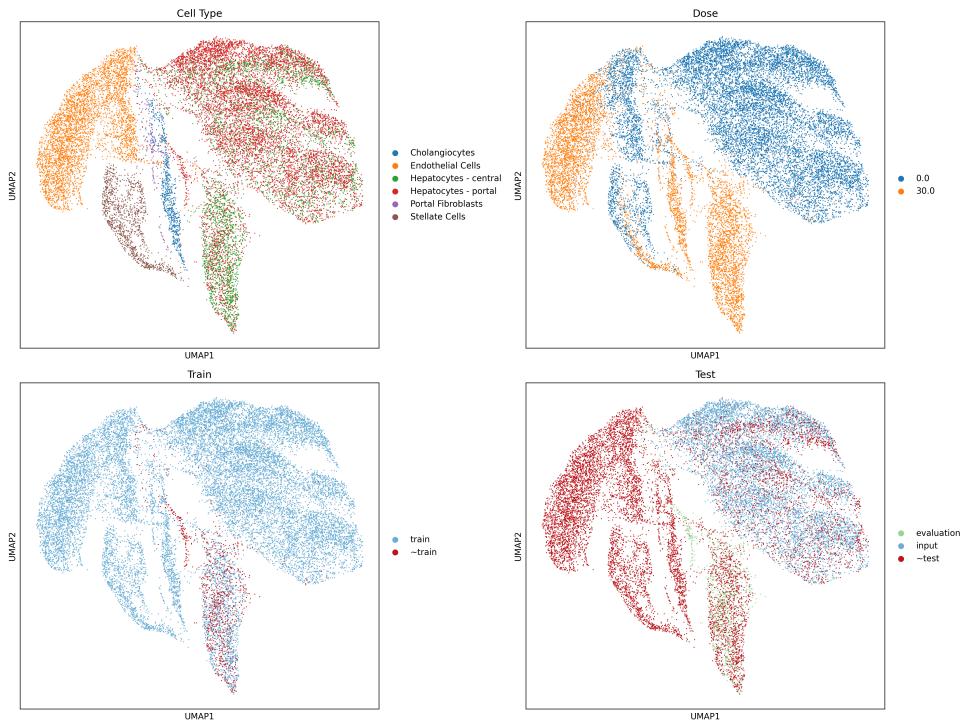
Figure 1.71: Wasserstein

1.3 Nault liver cell types evaluation

1.3.1 Multiple doses



1.3.2 Single dose 30 $\mu\text{g}/\text{kg}$



1.3.3 Comparison

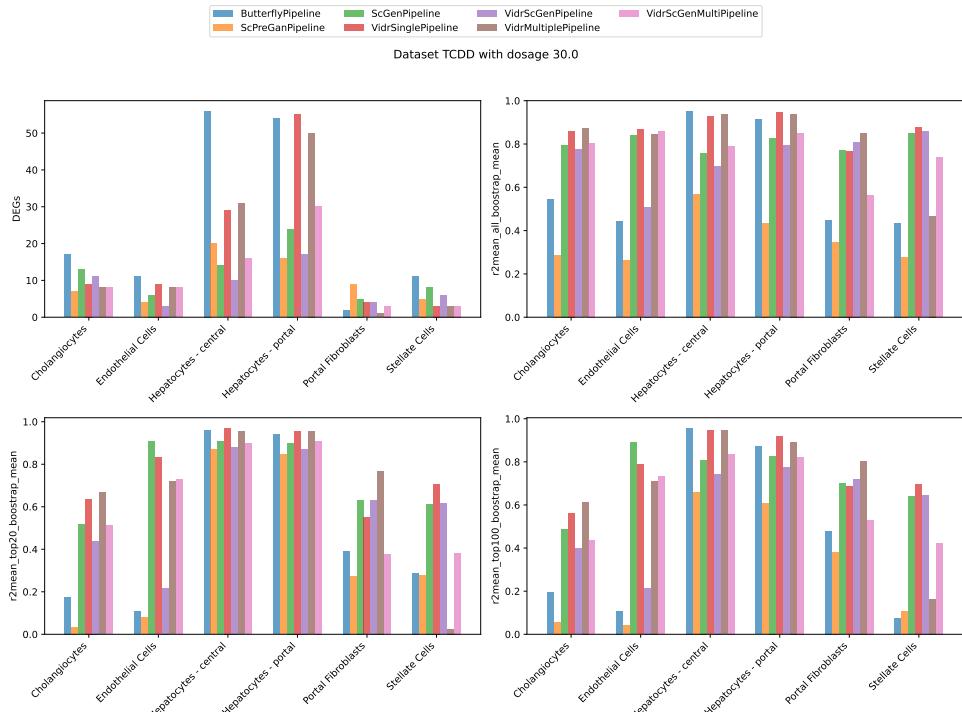


Figure 1.72: Baseline metrics for highest dosage 30 $\mu\text{g}/\text{kg}$ across cell types

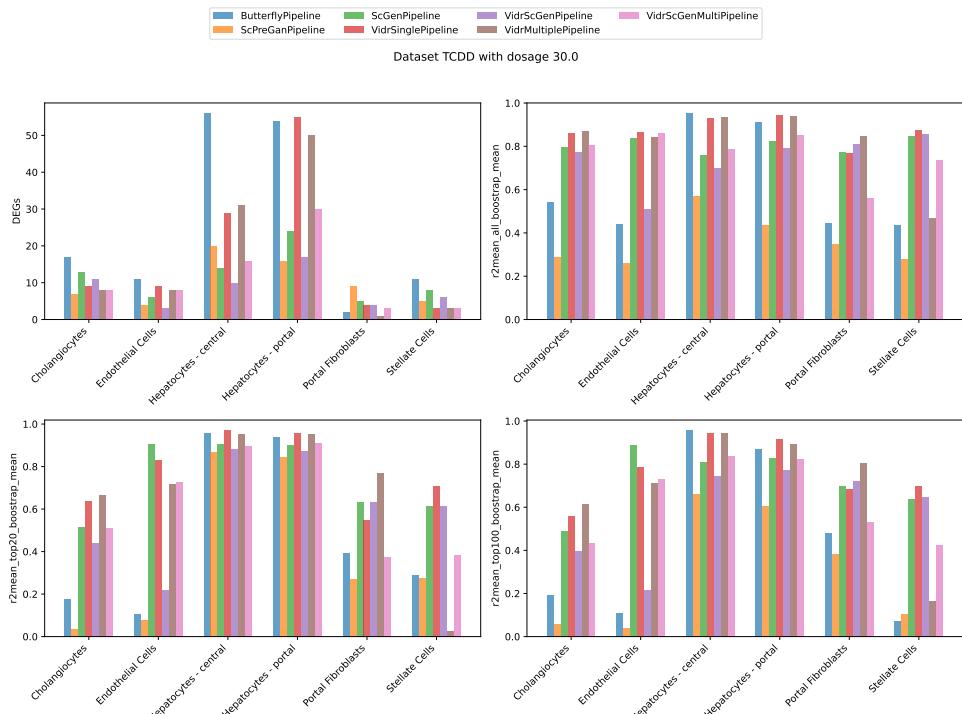


Figure 1.73: Distance metrics for highest dosage 30 $\mu\text{g}/\text{kg}$ across cell types

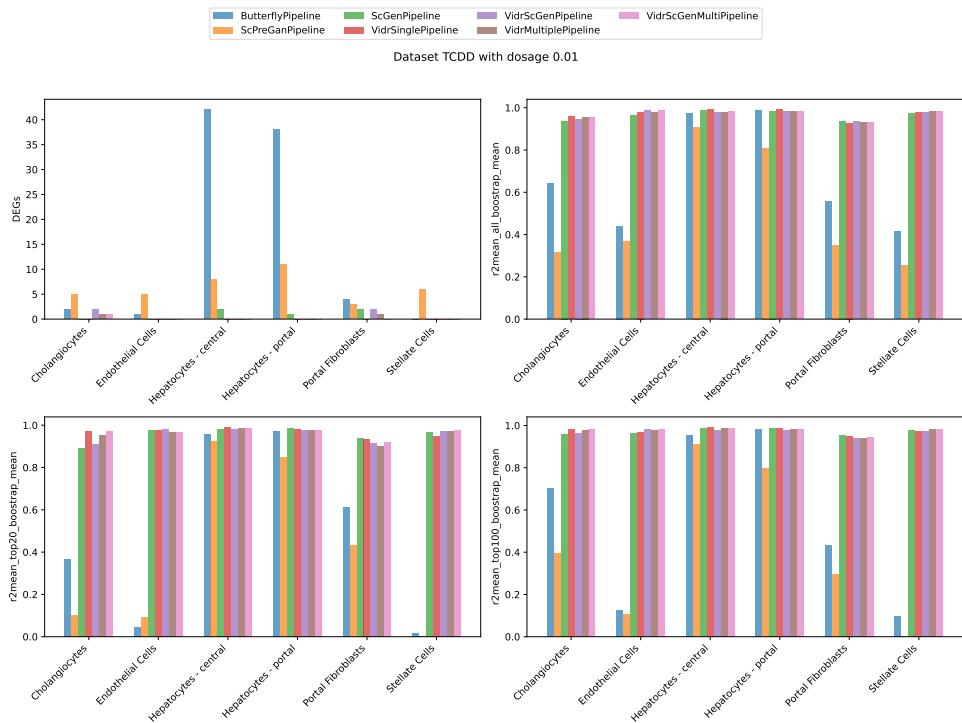


Figure 1.74: Baseline metrics for highest dosage $0.1\mu\text{g}/\text{kg}$ across cell types

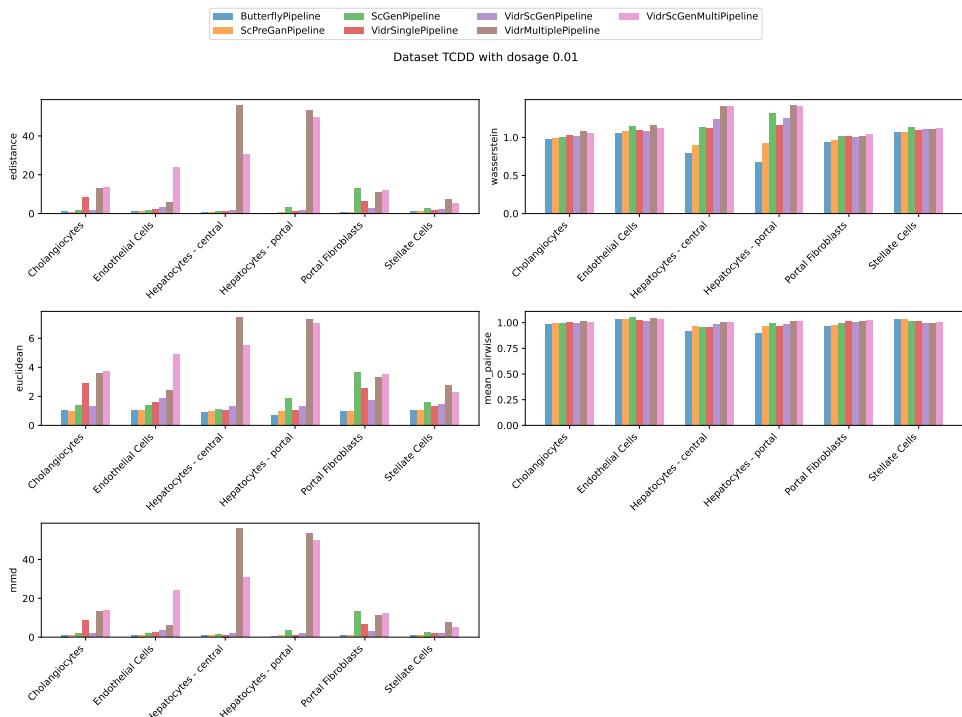


Figure 1.75: Distance metrics for highest dosage $0.1\mu\text{g}/\text{kg}$ across cell types

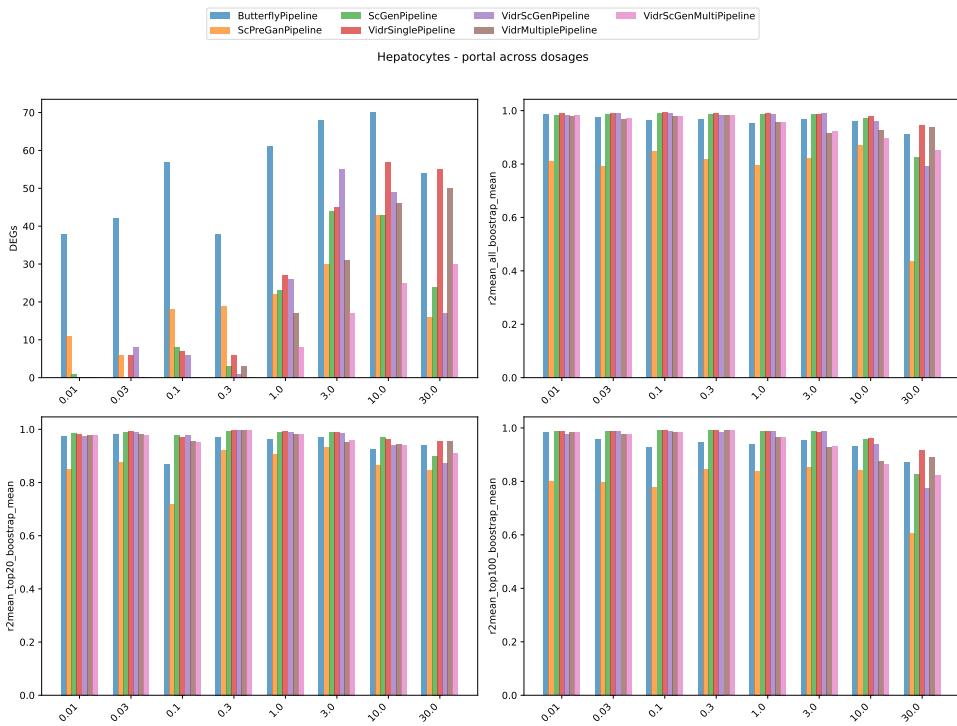


Figure 1.76: Baseline metrics for Hepatocytes - portal across dosages

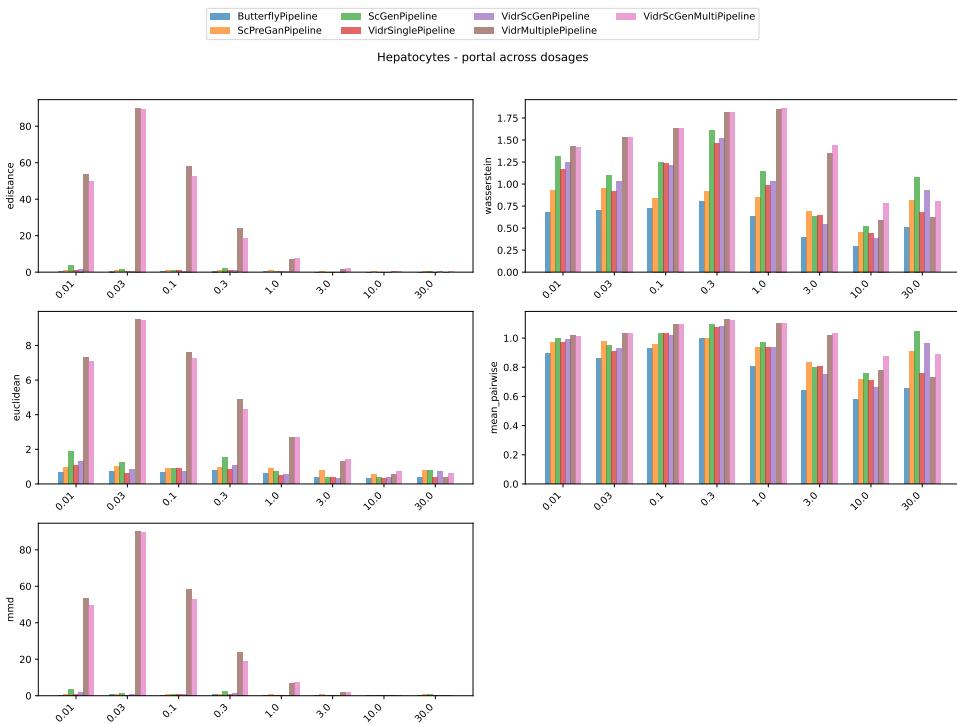


Figure 1.77: Distance metrics for Hepatocytes - portal across dosages

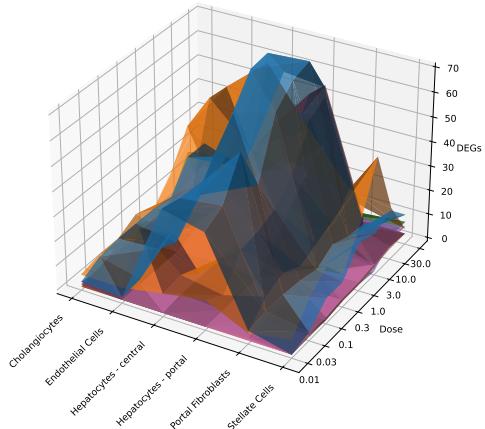
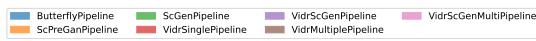


Figure 1.78

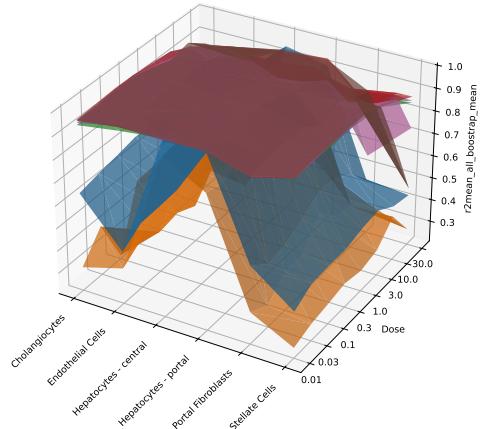


Figure 1.79

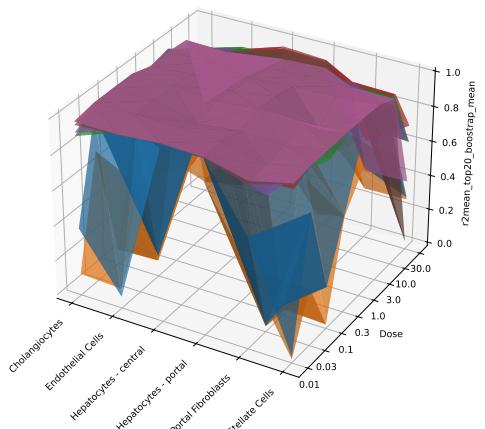


Figure 1.80

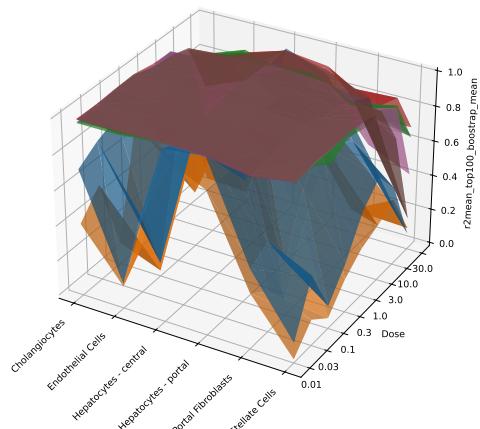


Figure 1.81

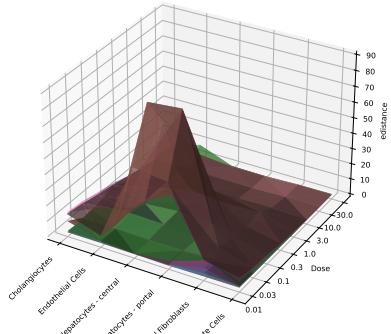


Figure 1.82: E-distance

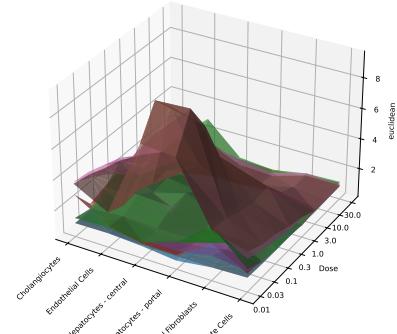


Figure 1.83: Euclidean

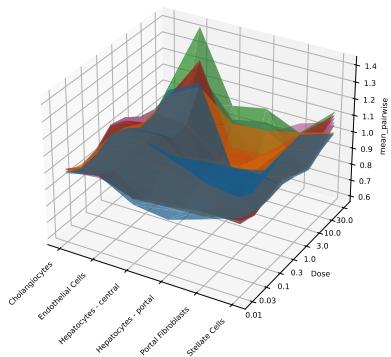


Figure 1.84: Mean pairwise

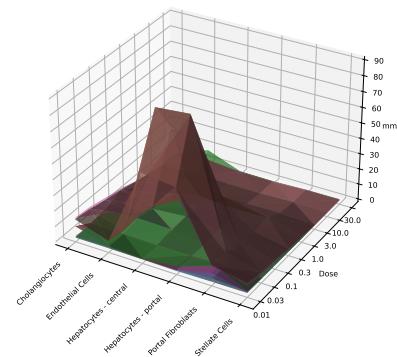


Figure 1.85: MMD

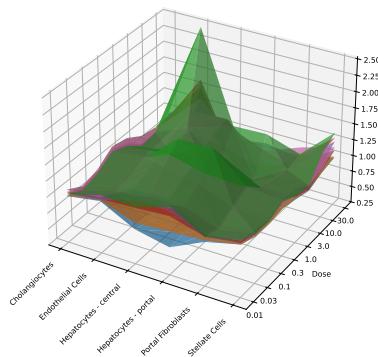
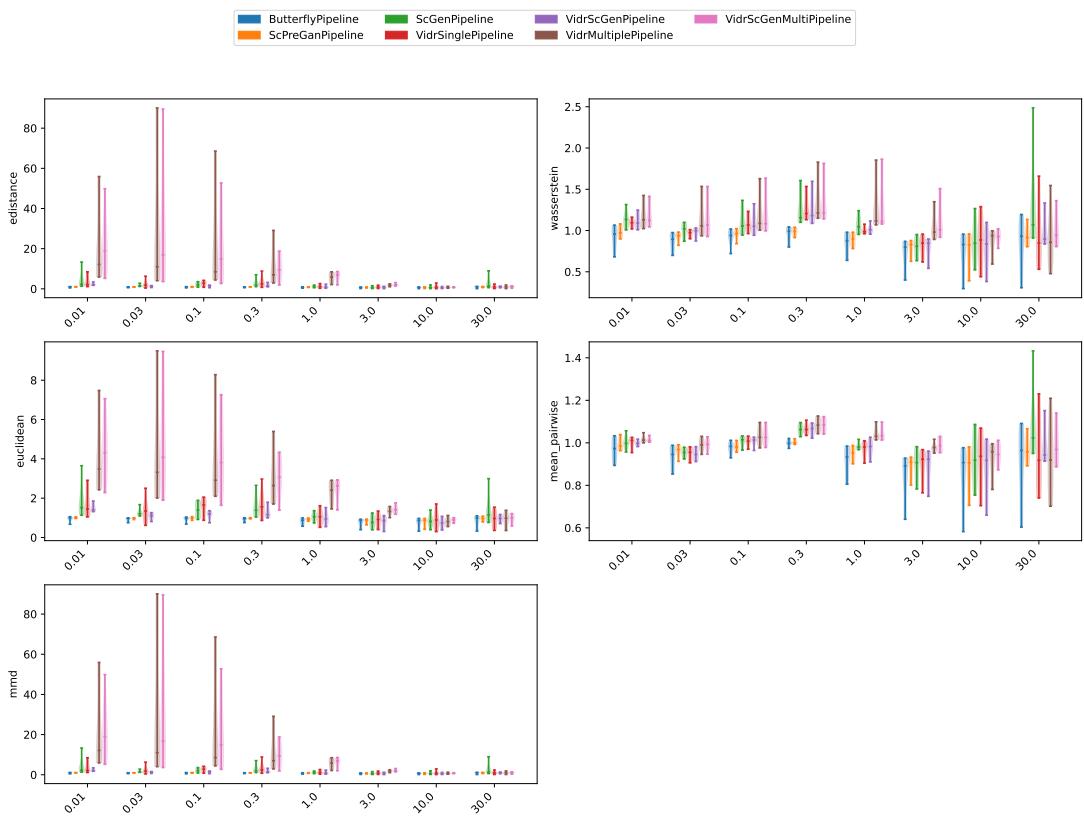
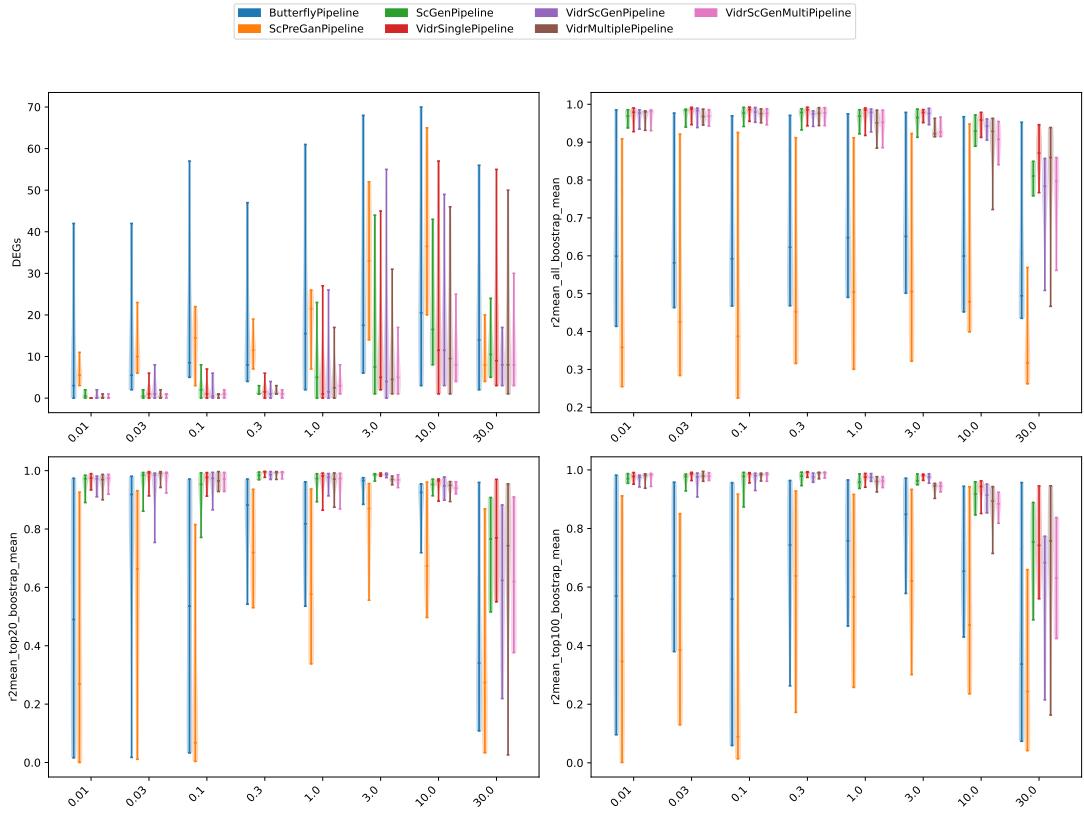
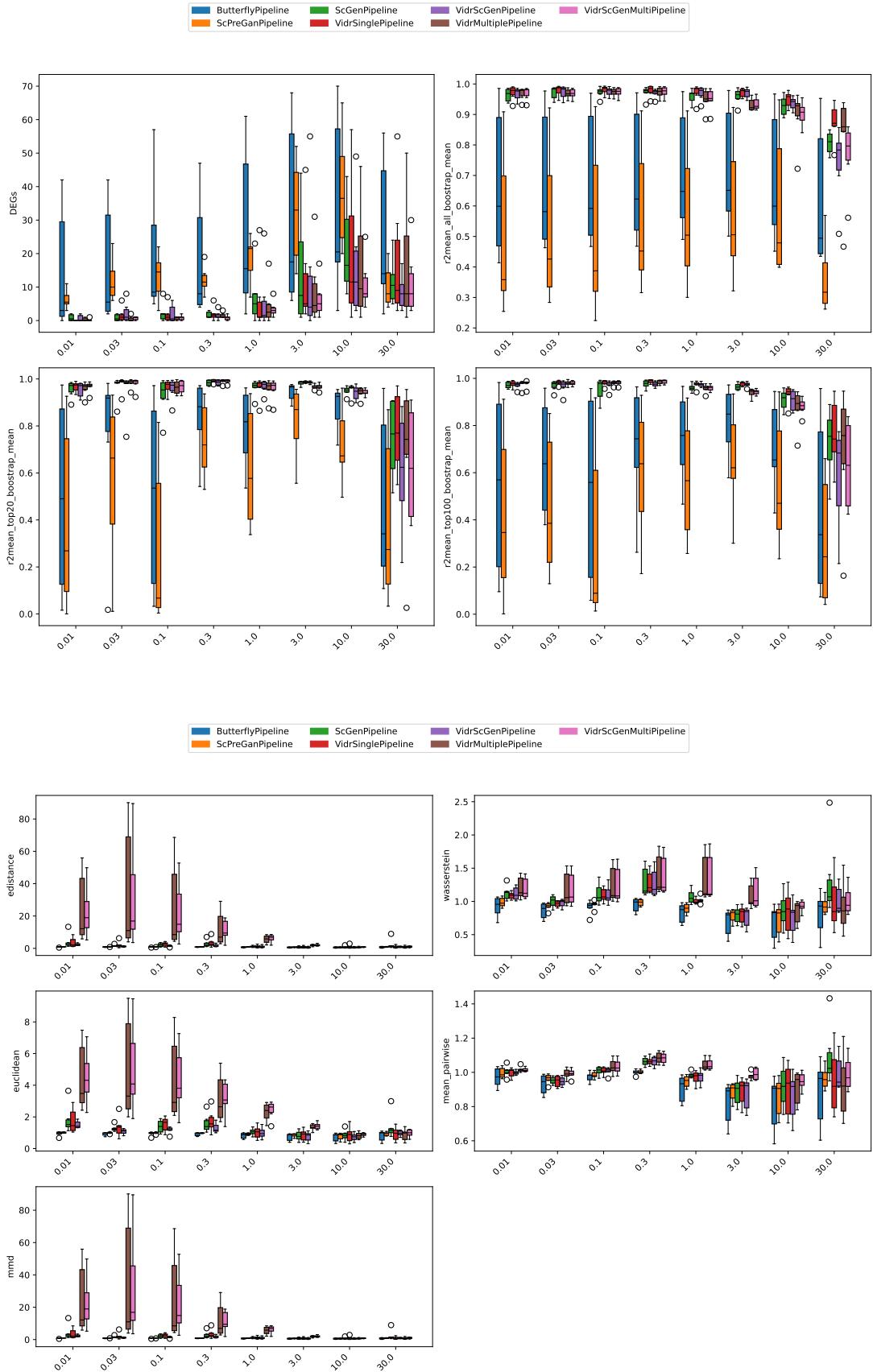
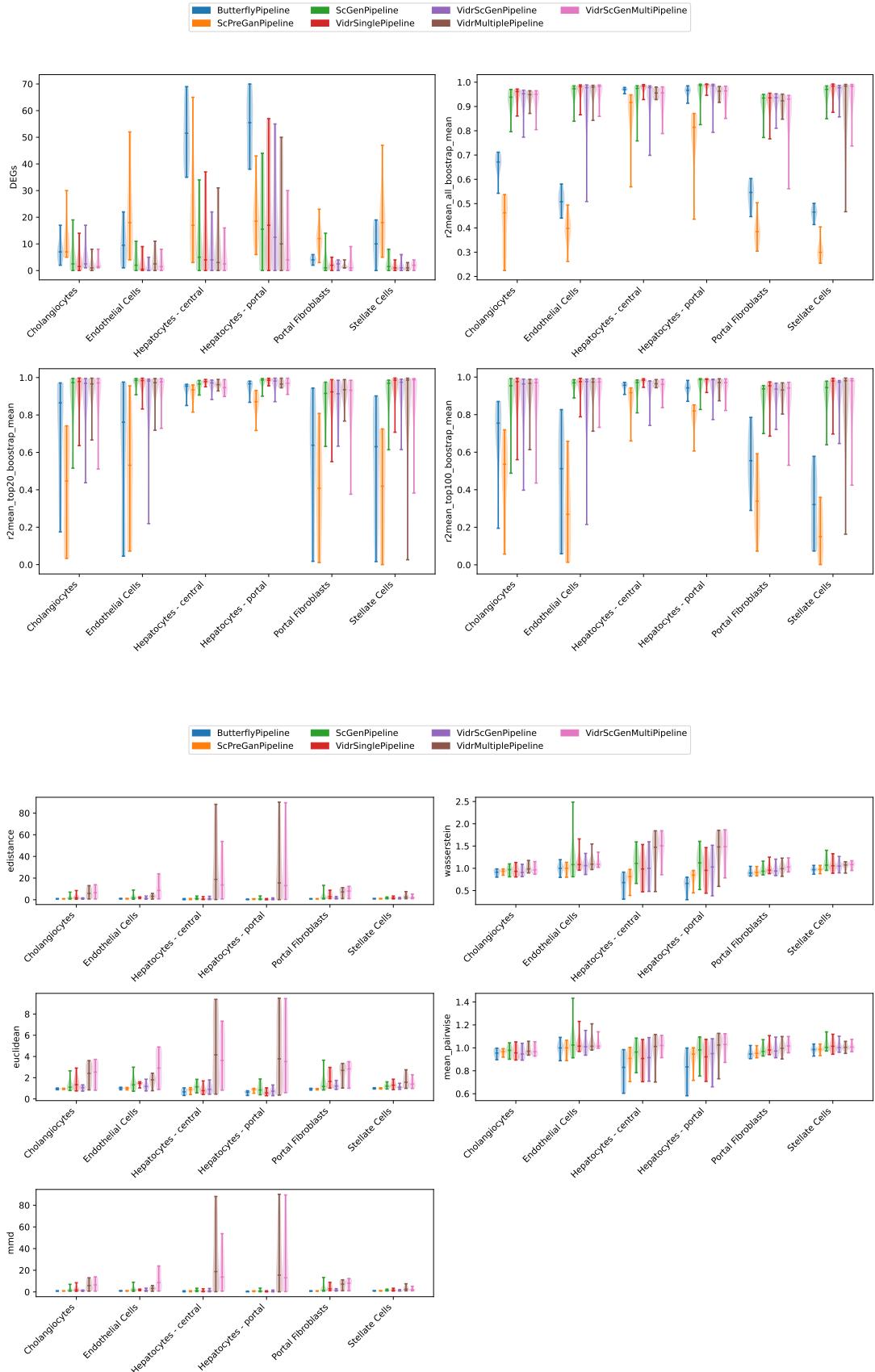


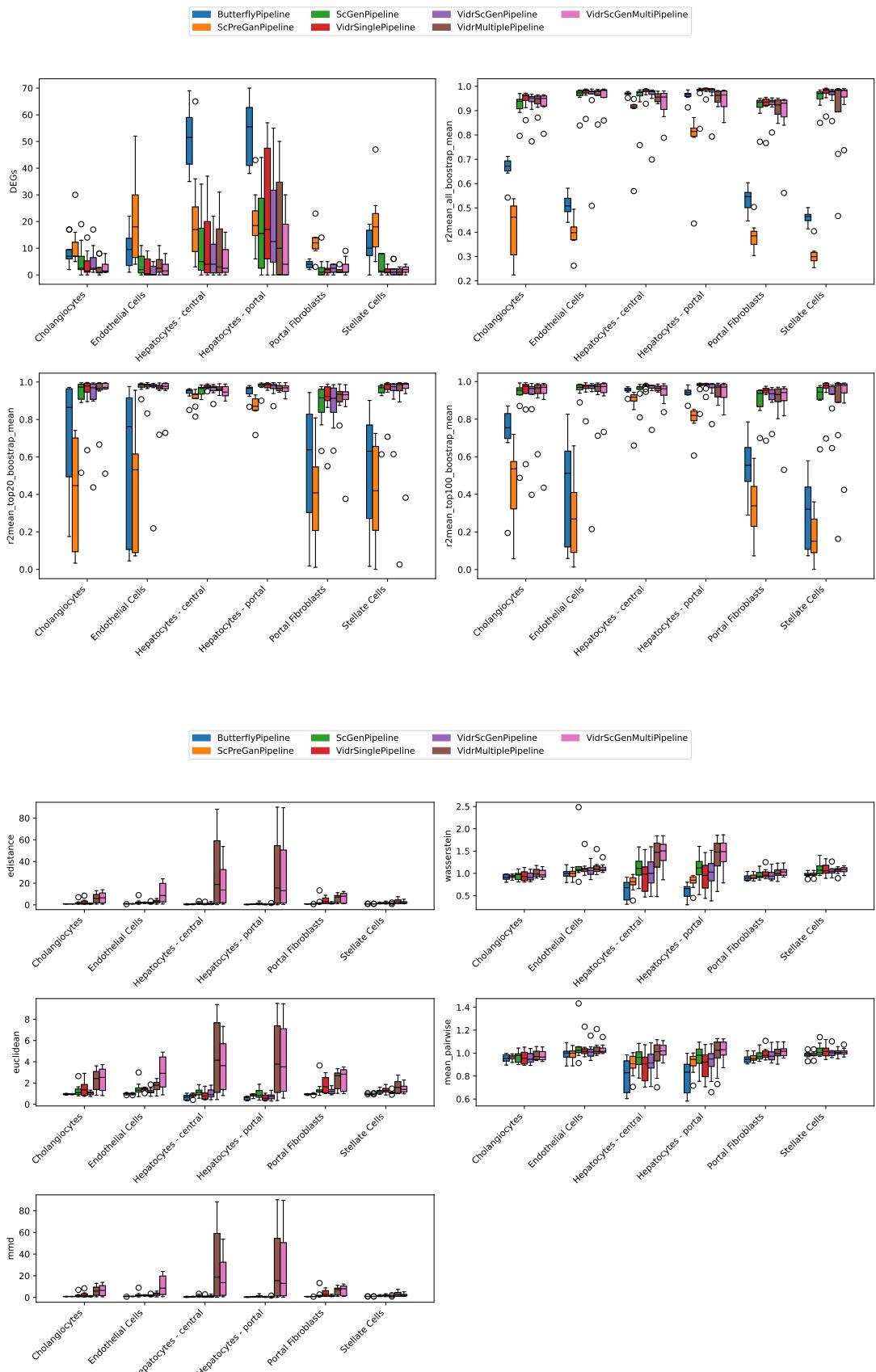
Figure 1.86: Wasserstein

Figure 1.87: Distance metrics per cell type









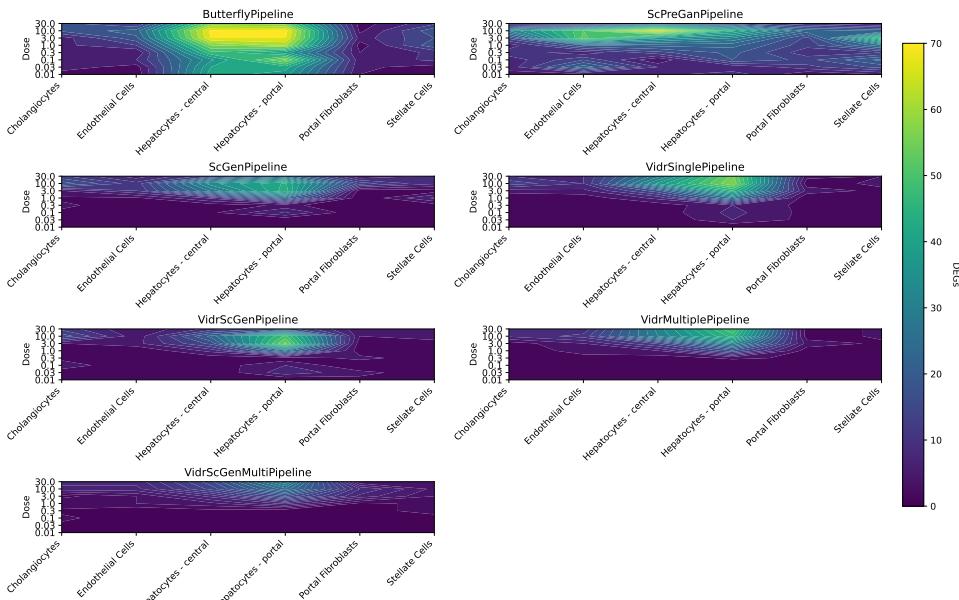


Figure 1.88: DEGs

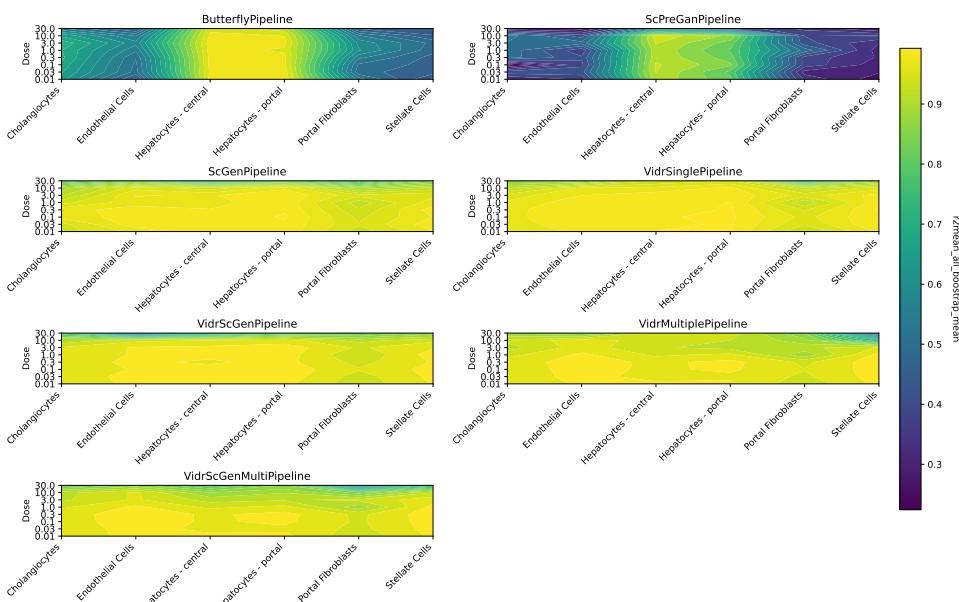


Figure 1.89: r² HVGs

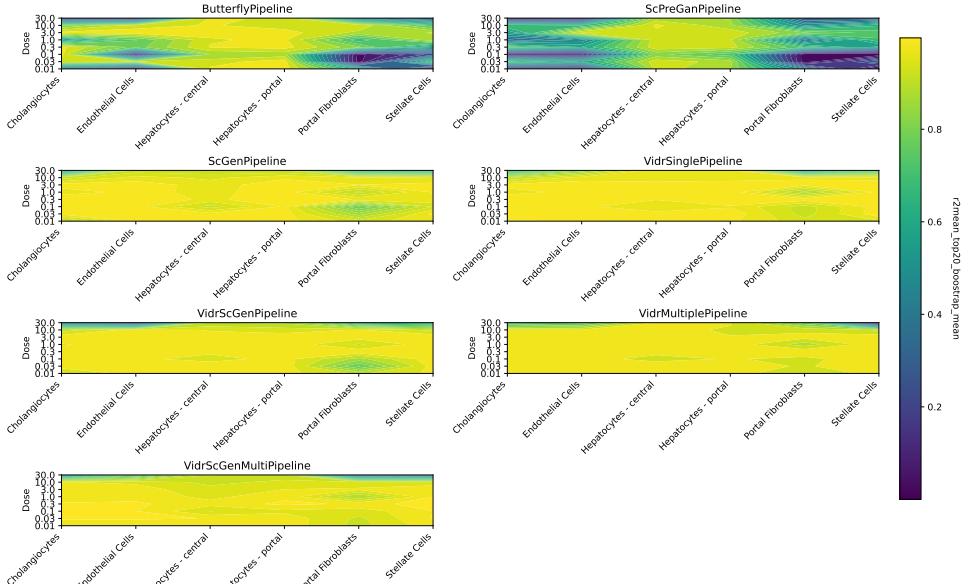


Figure 1.90: r^2 top 20

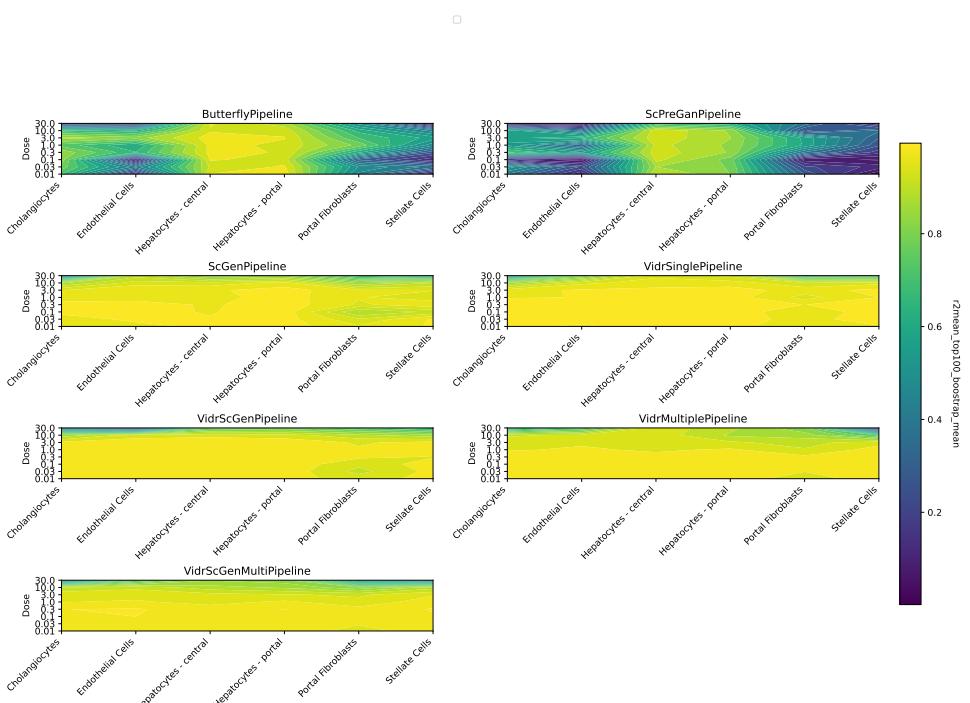


Figure 1.91: r^2 top 100

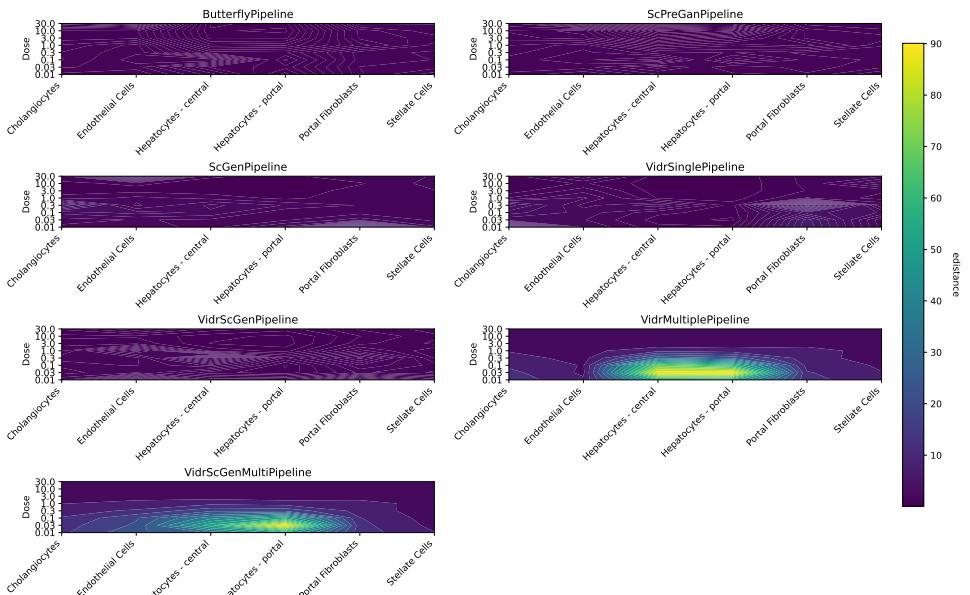
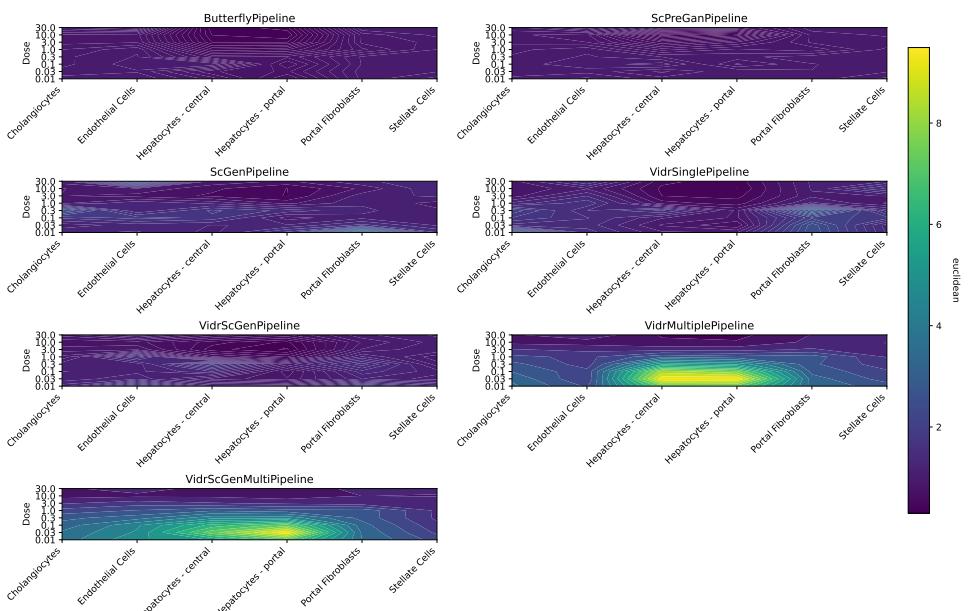


Figure 1.92: E-distance



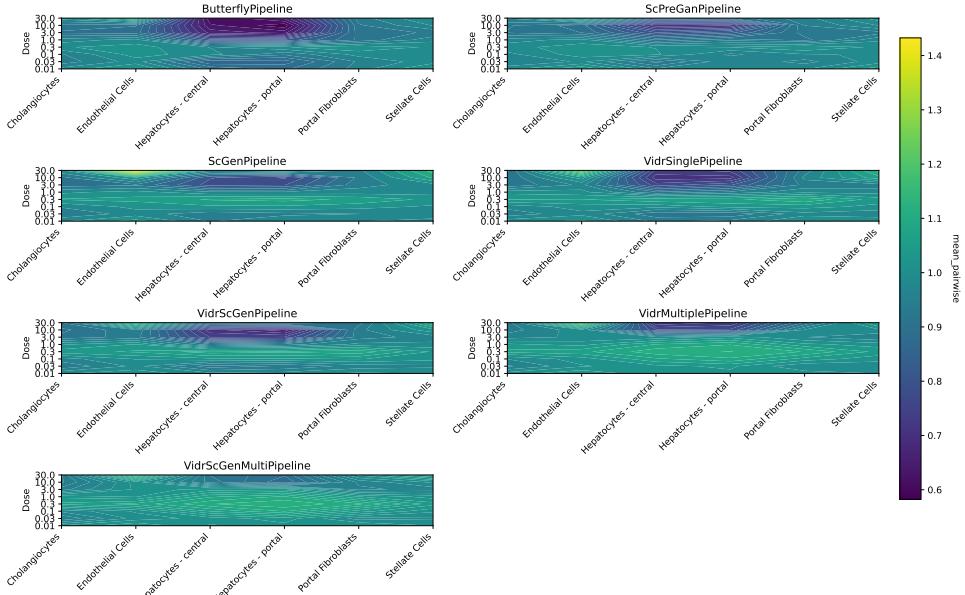


Figure 1.93: Mean pairwise

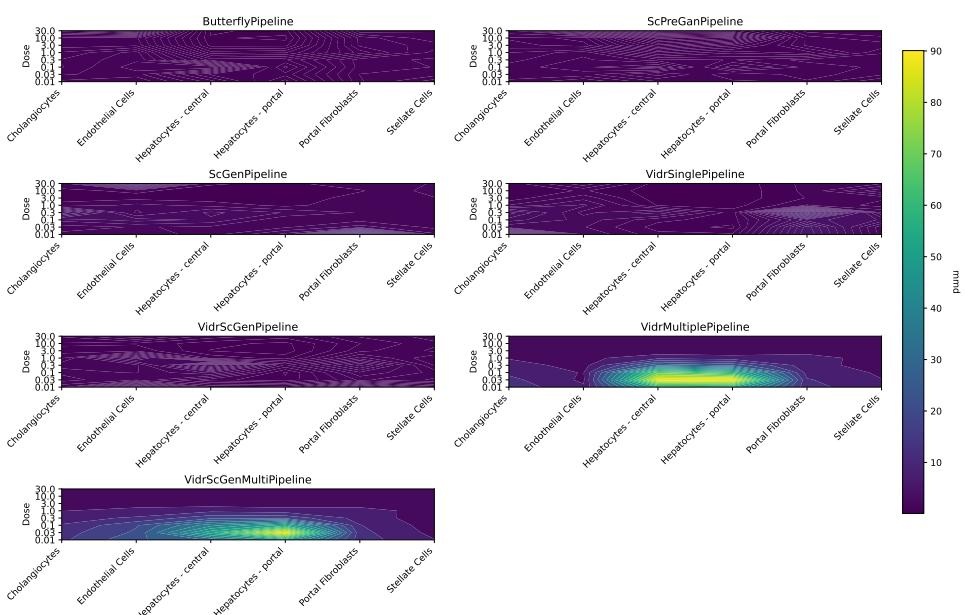


Figure 1.94: MMD

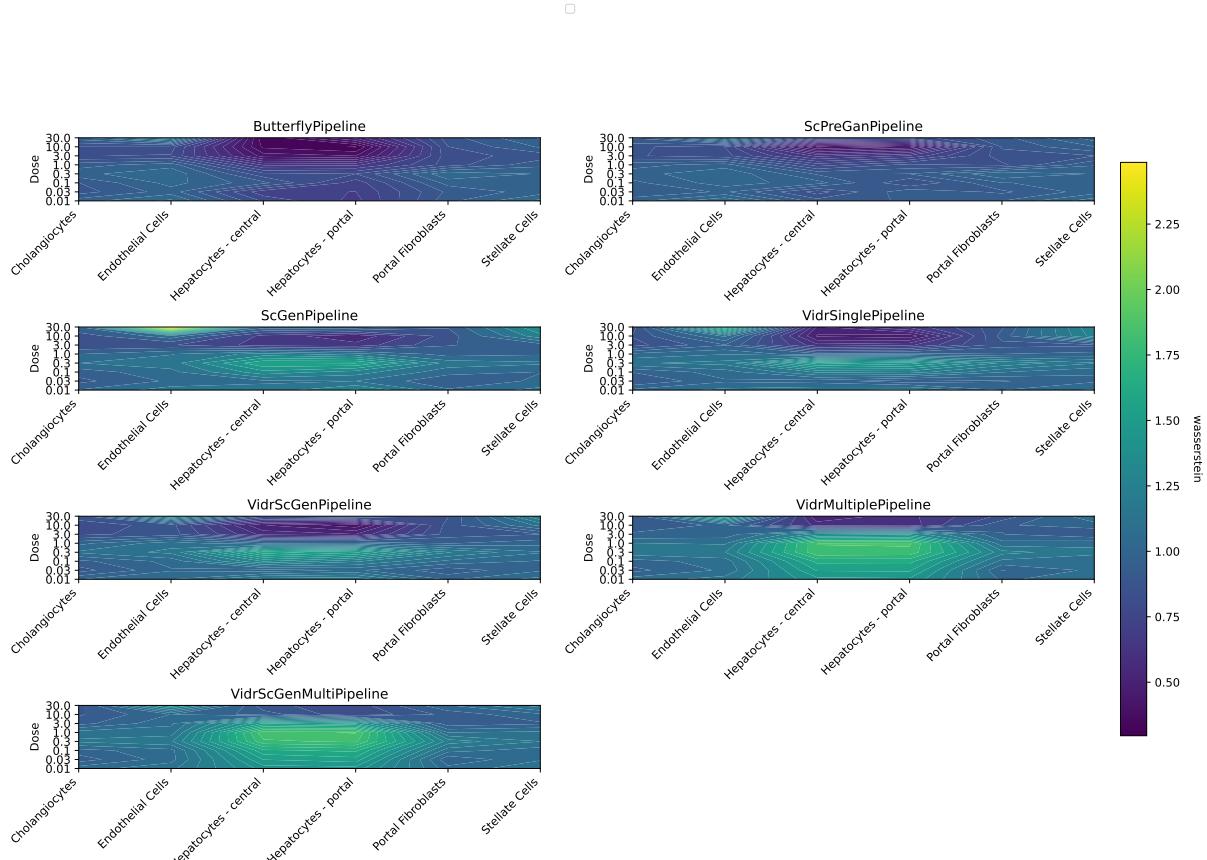
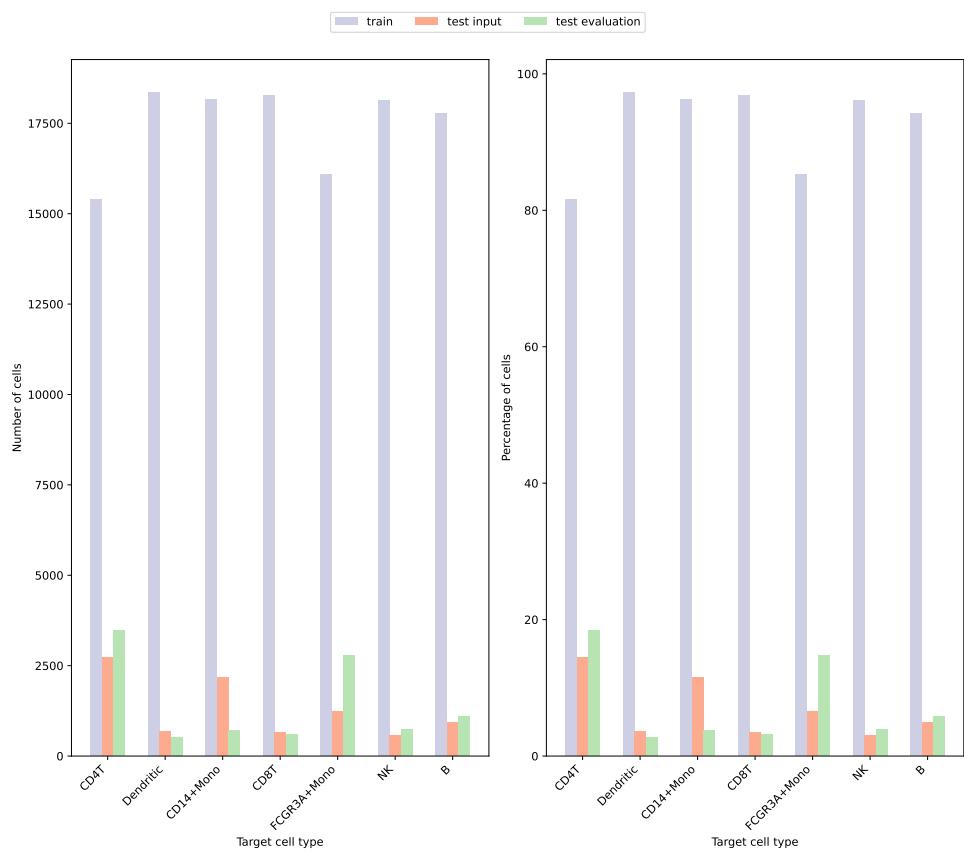
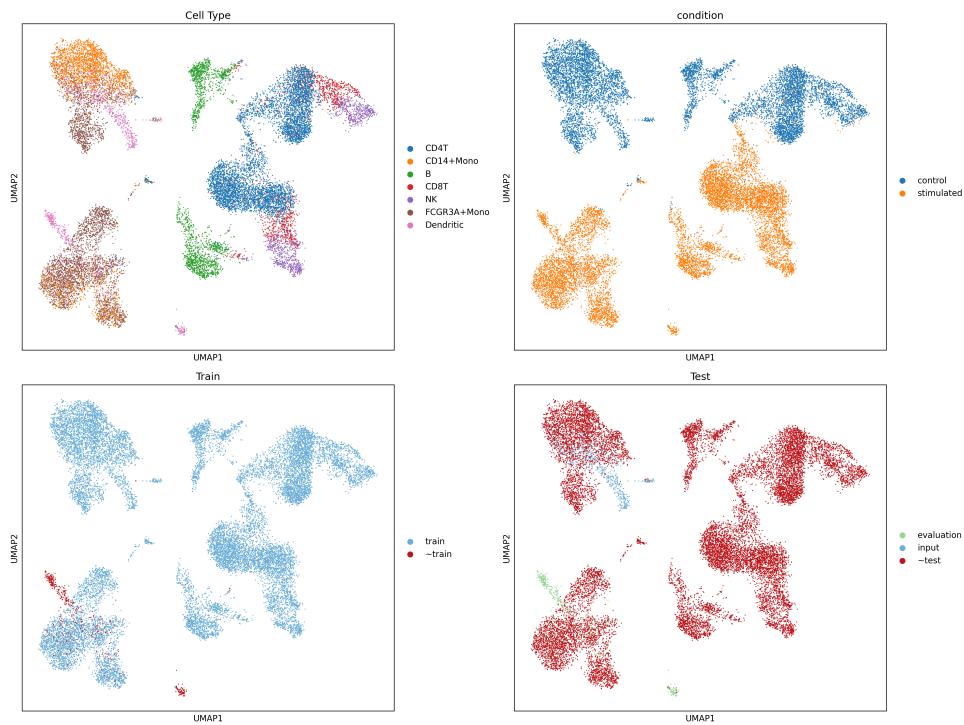


Figure 1.95: Wasserstein

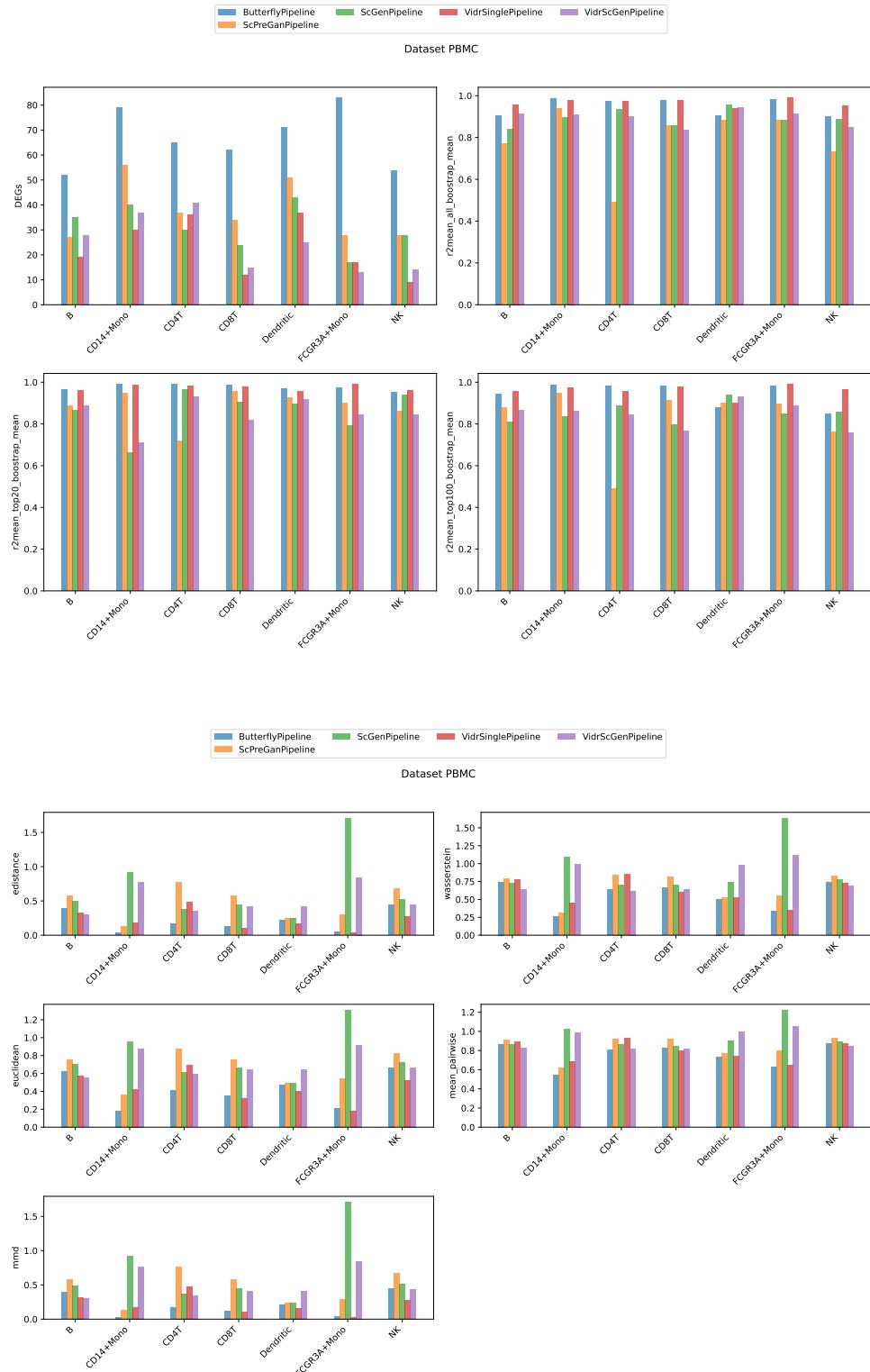
1.3.4 Παρατηρήσεις

- Το scButterfly και το scPreGan έχουν παρόμοια συμπεριφορά στις μετρικές και εμφανίζουν μεγάλη διακύμανση κατά μήκος των τύπων των κυττάρων και των δόσεων.
- Τα μοντέλα που έχουν ως βάση την αρχιτεκτονική του scGen (scVIDR, και οι παραλλαγές του), VAR και post-processing στο latent space, έχουν την υψηλότερη και πιο σταθερή απόδοση σε μετρικές του R^2 , ωστόσο υστερούν στην καταμέτρηση των κοινών διαφοροποιήσιμων γονιδίων έκφρασης (DEGs).

1.4 PBMC



1.4.1 Comparison



Appendix A

Ακρωνύμια και συντομογραφίες

LAN Local Area Network

Bibliography

[1] Google.