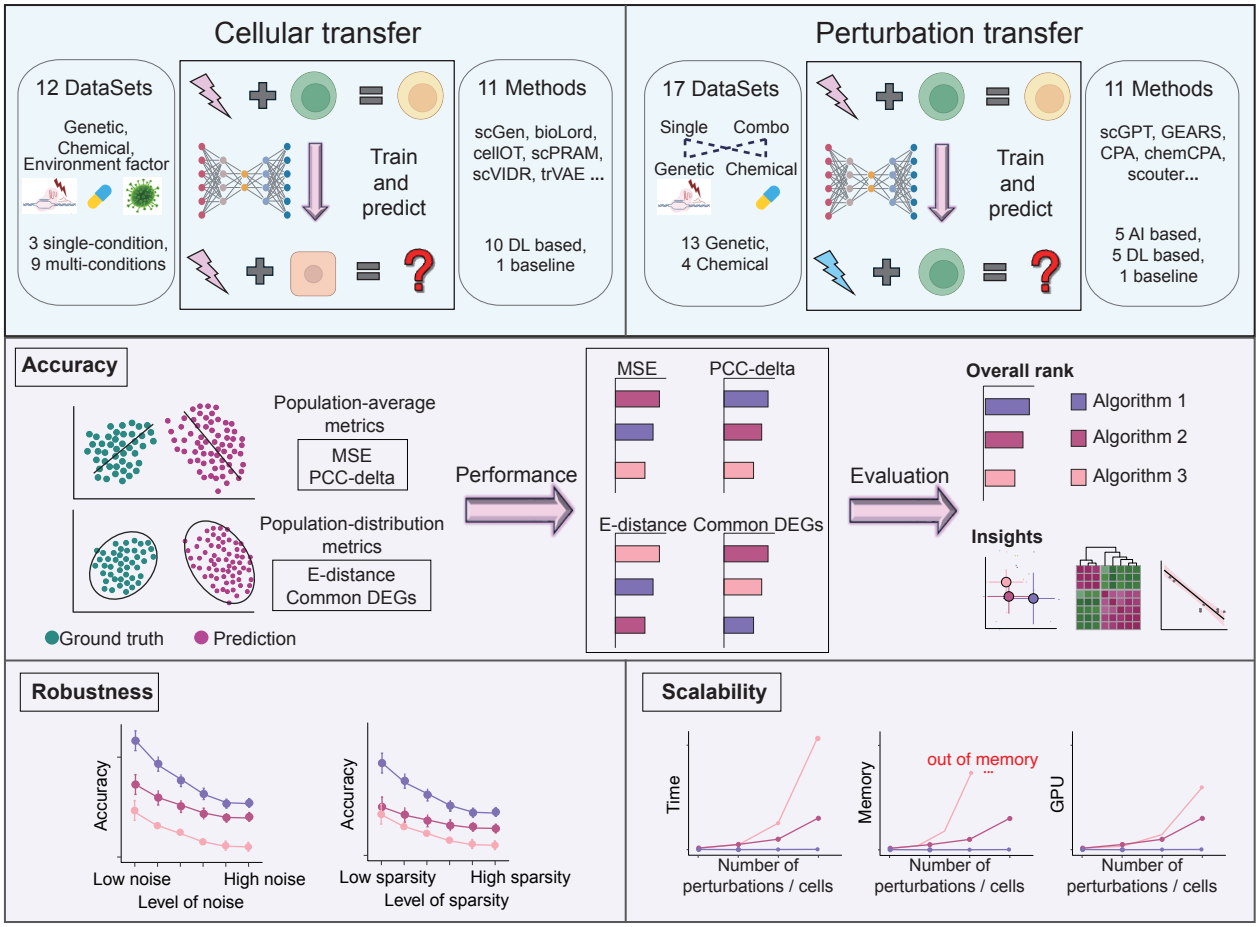
a



b

Dataset ID/name	Metadata	Contexts	Conditions	Context counts	Perturbation counts	Cell counts
1: kangCrossCell	None	Cell lines	Single	8	2	13,576
2: Parekh		Cell lines	Single	3	11	4,346
3: kangCrossPatient		Patients	Single	8	2	13,093
4: Haber	Batch	Cell lines	Multi	8	4	9,842
5: KaggleCrossPatient		Patients	Multi	3	11	25,583
6: crossPatient		Patients	Multi	6	3	117,363
7: KaggleCrossCell		Cell lines	Multi	5	11	23,653
8: crossSpecies	Time-point	Species	Multi	4	2	112,903
9: Afriat		Cell lines	Multi	2	3	19,053
10: McFarland		Cell lines	Multi	5	12	8,464
11: TCDD	Dosage	Cell lines	Multi	6	2	103,745
12: Sciplex3		Cell lines	Multi	3	10	26,046

c

Dataset ID/name	Perturbation type	Is Combination	Perturbation counts	Cell counts
1: Adamson	Genetic	No	77	56,998
2: Frangieh		No	212	110,188
3: TianActivation		No	93	20,169
4: TianInhibition		No	181	31,815
5: Replogle-exp7		No	105	102,148
6: Replogle-exp8		No	62	54,956
7: Papalexi		No	26	20,343
8: Replogle-RPE1essential		No	1331	173,912
9: Replogle-K562essential		No	1618	263,245
10: Norman		Yes	227	96,994
11: Wessels		Yes	128	19,775
12: Schmidt		Yes	148	60,572
13: Replogle-exp6		Yes	70	27,104
14: sciplex3-A549	Chemical	No	188	82,975
15: sciplex3-MCF7		No	188	223,630
16: sciplex3-K562		No	188	150,013
17: sciplex3-comb		Yes	32	63,378