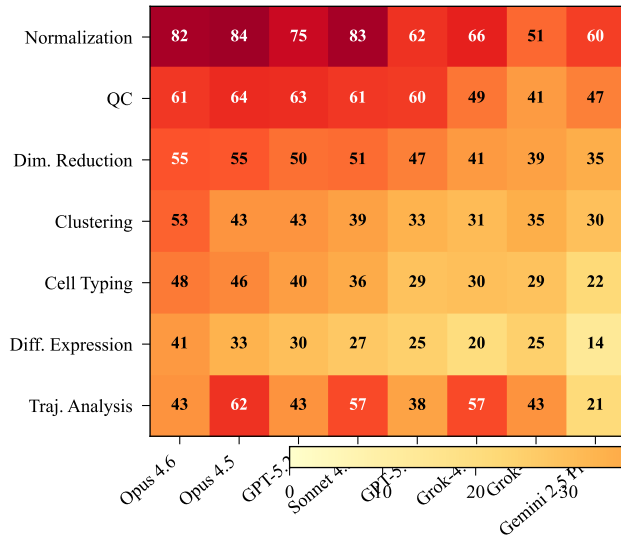(a) By Task Category | (b) By Platform

Heatmaps showing Accuracy (%) by model.

(a) By Task Category

	Opus 4.6	Opus 4.5	GPT-5.2	Sonnet 4.	GPT-5.0	Grok-4.20	Grok-...	Gemini 2.5 Pr
Normalization	82	84	75	83	62	66	51	60
QC	61	64	63	61	60	49	41	47
Dim. Reduction	55	55	50	51	47	41	39	35
Clustering	53	43	43	39	33	31	35	30
Cell Typing	48	46	40	36	29	30	29	22
Diff. Expression	41	33	30	27	25	20	25	14
Traj. Analysis	43	62	43	57	38	57	43	21

(b) By Platform

	Opus 4.	Opus 4.	GPT-5..	Sonnet 4.	GPT-5...	Grok-4	Grok-4	Gemini 2.5 Pro
CSGenetics	75	77	65	71	47	46	40	52
BD Rhapsody	53	56	55	54	42	43	30	31
Illumina	52	51	54	41	50	43	39	34
Chromium	52	47	47	46	46	42	41	33
ParseBio	53	42	36	33	25	25	32	26
MissionBio	42	38	23	34	20	19	25	10

Accuracy (%): 30 40 50 60 70 80 90