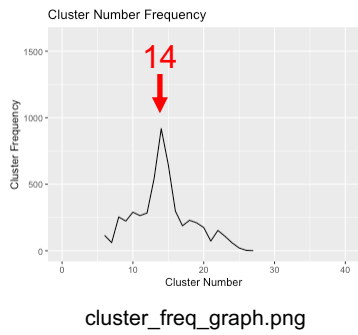


Figure S2

A



B

cluster_nu	freq	dim	k_param
14	22	36	35
14	22	36	34
14	22	36	32
14	21	36	33
14	21	36	31
14	20	37	36
14	20	36	36

cluster_number_matrix_sorted.csv

C

		Resolution parameter range									
		0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2	2.2
Prune parameter range	0.02	6	8	9	10	11	12	13	13	14	14
	0.04	6	8	10	11	12	13	14	14	14	15
	0.06	7	9	10	11	13	14	14	14	14	16
	0.08	8	10	11	14	14	14	14	14	17	18
	0.1	8	11	13	14	14	14	15	17	18	18
	0.12	8	12	13	14	14	15	16	18	19	20
	0.14	11	13	14	15	17	17	20	20	22	23
	0.16	13	14	14	16	17	20	20	21	23	24

res_prune_matrix_dims_36_k_35.csv

Figure 2. Method for defining the most robust PC dimensions, k-parameter, prune and resolution values. A) Graph (cluster_freq_graph.png) showing frequency distribution of clustering solutions. The most frequent clustering solution in these 5,120 iterations was 14 clusters, indicated by the red arrow. B) All cluster solutions and their frequency are listed in the cluster_number_matrix_sorted.csv table. The highest frequency that produced 14 clusters is shown in (B). The red box indicates the PC dimensions (dim = 36) and k-parameters (k_param = 35) that produced 14 clusters with the highest frequency. These were, therefore, chosen for downstream analysis. All clustering solutions using PC dims = 36 and k_param = 35 when using a range of prune and resolution values are listed in res_prune_matrix_dims_36_k_35.csv table (C). The red boxes indicate the final prune value (0.08) and resolution value (1.4) used for downstream analysis.