Supplementary Information:
Deep learning and automated Cell Painting reveal Parkinson's disease-specific signatures in primary patient fibroblasts

Lauren Schiff[1,3], Bianca Migliori[2,3], Ye Chen[1,3], Deidre Carter[2,3], Caitlyn Bonilla[1], Jenna Hall[2], Minjie Fan[1], Edmund Tam[2], Sara Ahadi[1], Brodie Fischbacher[2], Anton Geraschenko[1], Christopher J. Hunter[2], Subhashini Venugopalan[1], Sean DesMarteau[2], Arunachalam Narayanaswamy[1], Selwyn Jacob[2], Zan Armstrong[1], Peter Ferrarotto[2], Brian Williams[1], Geoff Buckley-Herd[2], Jon Hazard[1], Jordan Goldberg[2], Marc Coram[1], Reid Otto[2], Edward A. Baltz[1], Laura Andres-Martin[2], Orion Pritchard[1], Alyssa Duren-Lubanski[2], Kathryn Reggio[2], NYSCF Global Stem Cell Array Team[2], Lauren Bauer[2], Raeka S. Aiyar[2], Elizabeth Schwarzbach[2], Daniel Paull[2], Scott A. Noggle[2], Frederick J. Monsma, Jr.[2], Marc Berndl[1,4], Samuel J. Yang[1,4], Bjarki Johannesson[2,4]

[1]Google Research, Mountain View, California, USA. [2]The New York Stem Cell Foundation Research Institute, New York, New York, USA. [3]These authors contributed equally: Lauren Schiff, Bianca Migliori, Ye Chen, Deidre Carter. [4]These authors contributed equally: Marc Berndl, Samuel J. Yang, Bjarki Johannesson. e-mail: marcberndl@google.com; samuely@google.com; bjohannesson@nyscf.org

Cell line ID	Donor ID	Cross-val fold	Pair ID	Disease state	Sex	Age	European ancestry	UPDRS score	Biopsy collection year	Biopsy location	Thaw format	Thaw freeze date	Doubling time
01	51211	1	0	Healthy	F	54	99%		2012	unspecified	6w	6/17/2019	3.32
02	51255	1	0	*LRRK2* PD	F	56	92%	71	2017	left upper thigh	6w	6/17/2019	2.66
03	51260	2	1	Healthy	F	64	98%		2017	left upper leg	12w→6w	8/9/2019	2.70
04	51253	2	1	Sporadic PD	F	63	99%	23	2017	left upper arm	6w	7/30/2019	3.41
05	50114	4	2	Healthy	M	67	94%		2012	unspecified	6w	7/31/2019	2.77
06	50463	4	2	Sporadic PD	M	69	92%	34	2013	unspecified	6w	8/15/2019	3.39
07	50752	4	3	Healthy	F	58	90%		2013	left upper arm	12w→6w	8/22/2019	5.36
08	50437	4	3	Sporadic PD	F	59	90%	45	2013	left upper leg	6w	7/11/2019	3.80
09	50665	1	4	Healthy	F	71	91%		2013	right lower leg	12w→6w	7/30/2019	3.24
10	50373	1	4	Sporadic PD	F	69	100%	61	2012	right thigh	6w	7/5/2019	4.37
11	51218	3	5	Healthy	F	60	100%		2016	left upper arm	6w	7/11/2019	2.47
12	51193	3	5	Sporadic PD	F	60	91%	53	2015	left thigh	6w	7/22/2019	2.49
13	50584	4	6	Healthy	M	81			2013	right lower cheek	6w	6/7/2019	3.11
14	51259	4	6	Sporadic PD	M	76	95%	29	2017	left arm	6w	9/3/2019	3.87
15	51005	4	7	Healthy	F	57	90%		2014	left upper arm	6w	7/5/2019	2.83
16	51149	4	7	Sporadic PD	F	57	97%	26	2015	left upper arm	6w	7/5/2019	2.50
17	50183	3	8	Healthy	M	62	98%		2012	chest	6w	7/18/2019	6.19
18	50951	3	8	Sporadic PD	M	62	98%	20	2014	left upper arm	12w→6w	7/30/2019	3.30
19	51107	0	9	Healthy	M	59	97%		2014	upper arm	6w	7/11/2019	2.53
20	51099	0	9	Sporadic PD	M	61	100%	19	2014	upper left arm	12w→6w	7/22/2019	4.06
21	51183	1	10	Healthy	M	66	89%		2015	left arm	6w	8/9/2019	2.12
22	50480	1	10	Sporadic PD	M	66	99%	56	2013	left upper arm	6w	7/31/2019	4.55
23	50764		11	Healthy	F	56	89%		2013	left upper arm	6w	6/7/2019	2.71
24	51266		11	*GBA* PD	F	54	96%	30	2017	unspecified	6w	6/7/2019	2.28
25	50167	0	12	Healthy	M	62	98%		2012	unspecified	12w→6w	7/22/2019	4.64
26	51156	0	12	Sporadic PD	M	68	100%	20	2015	left upper arm	6w	8/13/2019	2.41
27	50956		13	Healthy	M	47			2014	left upper arm	6w	6/11/2019	6.64
28	51249		13	*GBA* PD	M	46	92%	20	2016	right upper arm	6w	6/11/2019	2.44
29	50767	2	14	Healthy	F	71	88%		2013	left upper arm	6w	7/30/2019	2.71
30	50406	2	14	Sporadic PD	F	68	90%	24	2012	right upper arm	12w→6w	9/3/2019	5.24
31	51105	2	15	Healthy	M	59			2014	upper arm	6w	6/19/2019	2.28
32	51142	2	15	Sporadic PD	M	59	90%	36	2015	right arm	12w→6w	7/5/2019	3.07
33	51140	0	16	Healthy	M	55	96%		2015	left upper arm	6w	6/13/2019	3.53
34	51256	0	16	Sporadic PD	M	54	94%	31	2017	left inner arm	6w	6/13/2019	6.57
35	50939	1	17	Healthy	M	64	91%		2014	left upper arm	6w	7/22/2019	4.36
36	50449	1	17	Sporadic PD	M	64	98%	44	2013	left upper arm	12w→6w	7/30/2019	3.41
37	50128	3	18	Healthy	M	57	97%		2012	unspecified	6w	7/11/2019	2.72
38	50392	3	18	*LRRK2* PD	M	59	93%	20	2012	right thigh	6w	7/11/2019	3.58
39	51239	0	19	Healthy	M	64	99%		2016	left upper inner arm	6w	7/31/2019	2.25
40	51037	0	19	Sporadic PD	M	68	93%	34	2014	left upper arm	12w→6w	8/26/2019	4.29
41	50199	1	20	Healthy	M	81	99%		2012	right arm	6w	6/7/2019	3.03
42	50590	1	20	Sporadic PD	M	76	92%	27	2013	left upper arm	6w	8/13/2019	3.30
43	50112	0	21	Healthy	F	77	92%		2012	unspecified	12w→6w	7/5/2019	4.28
44	51126	0	21	Sporadic PD	F	74	99%	35	2015	unspecified	6w	6/17/2019	3.42
45	50192	2	22	Healthy	M	77	97%		2012	nose	6w	6/25/2019	4.53
46	51261	2	22	*LRRK2* PD	M	74	92%	29	2017	right upper arm	12w→6w	8/22/2019	2.78
47	51152		23	Healthy	F	74	100%		2015	left upper arm	12w→6w	6/19/2019	2.94
48*	10124*		23	*GBA* PD*	F	74	94%		2011	right arm	6w	6/13/2019	6.04
49	51030	2	24	Healthy	F	70	94%		2014	left arm	6w	7/18/2019	2.80
50	51250	2	24	Sporadic PD	F	69	98%	21	2016	left upper arm	6w	7/18/2019	3.22
51	51093	0	25	Healthy	F	73	99%		2015	left upper arm	12w→6w	7/11/2019	2.75
52	50864	0	25	Sporadic PD	F	73	98%	34	2013	left upper arm	6w	6/19/2019	3.48
53	51254		26	Healthy	M	66	92%		2017	unspecified	6w	7/5/2019	2.25
54	10198		26	*GBA* PD	M	66	98%	38	2012	right upper inner arm	6w	7/5/2019	2.40
55	51148	4	27	Healthy	M	61	90%		2015	left upper arm	6w	8/13/2019	2.68
56	50640	4	27	Sporadic PD	M	66	93%	21	2013	upper arm	6w	8/13/2019	2.90
57*	50634*	4	28	Healthy*	F	72	92%		2013	left upper arm	6w	7/5/2019	2.61
58	51243	4	28	Sporadic PD	F	72	90%	47	2016	right thigh	12w→6w	8/22/2019	2.66
59	51194	3	29	Healthy	M	80	99%		2015	left arm	6w	6/11/2019	3.75
60	51268	3	29	Sporadic PD	M	79	92%	44	2017	unspecified	6w	6/11/2019	2.70
61	51123	1	30	Healthy	F	54	92%		2015	left upper arm	6w	6/25/2019	3.21
62	50483	1	30	Sporadic PD	F	56	91%	42	2013	left upper leg	6w	6/25/2019	4.79
63	51004	3	31	Healthy	M	54	98%		2014	left upper arm	12w→6w	6/25/2019	3.09
64	50963	3	31	Sporadic PD	M	54	99%	24	2014	left upper arm	6w	6/7/2019	3.67
65	10130	4	32	Healthy	M	52			2011	left inner arm	6w	6/11/2019	2.52
66	50674	4	32	*LRRK2* PD	M	51	91%	12	2013	right upper arm	6w	6/11/2019	4.23
67	50598	2	33	Healthy	M	58	91%		2013	left forearm	6w	7/30/2019	2.57
68	50610	2	33	Sporadic PD	M	59	97%	28	2013	left upper arm	6w	6/25/2019	2.88
69	51162	1	34	Healthy	M	47	98%		2015	left arm	6w	8/13/2019	2.01
70	50492	1	34	*LRRK2* PD	M	45	91%		2013	left upper arm	12w→6w	8/22/2019	2.90
71	51235	0	35	Healthy	M	68	98%		2016	left arm	6w	8/9/2019	2.48
72	51212	0	35	Sporadic PD	M	70	98%	28	2015	left arm	6w	8/15/2019	3.08
73	50191		36	Healthy	M	61	92%		2012	right cheek	6w	7/22/2019	6.10
74	50660		36	*GBA* PD	M	65	91%		2013	left upper arm	12w→6w	7/30/2019	5.78
75	50105		37	Healthy	M	76	90%		2012	unspecified	12w→6w	7/22/2019	4.07
76	51221		37	*GBA* PD	M	74	93%	22	2016	left upper arm	6w	6/25/2019	2.63
77	51274		38	Healthy	M	64			2018	left upper arm	6w	8/13/2019	2.82
78	51021		38	*GBA* PD	M	65	93%		2014	left upper arm	6w	8/13/2019	2.86
79	50176		39	Healthy	M	58			2012	chin	6w	6/19/2019	3.16
80	50880		39	*GBA* PD	M	59	93%	18	2013	upper arm	6w	6/19/2019	3.50
81	50659	0	40	Healthy	F	64	91%		2013	left upper arm	6w	7/31/2019	2.58
82	51010	0	40	*LRRK2* PD	F	63	94%	18	2014	left upper arm	12w→6w	8/26/2019	3.91
83	50617	4	41	Healthy	M	53			2013	right upper leg	6w	6/13/2019	2.79
84	51176	4	41	Sporadic PD	M	53	98%	31	2015	left arm	6w	6/13/2019	2.71
85	51139	3	42	Healthy	F	52	99%		2015	left upper arm	6w	8/15/2019	2.38
86	51187	3	42	Sporadic PD	F	66	98%	16	2015	left upper arm	6w	8/26/2019	2.52
87	50758	3	43	Healthy	F	78	91%		2013	left upper arm	6w	6/13/2019	4.02
88	51200	3	43	Sporadic PD	F	80	90%	79	2015	left arm	6w	6/11/2019	4.54
89	50174	2	44	Healthy	M	56	98%		2012	forehead	6w	6/17/2019	2.55
90	50421	2	44	Sporadic PD	M	55	99%	26	2012	left upper arm	6w	6/17/2019	4.93
91	50437			Sporadic PD	F	64	90%	45	2019	left upper arm	6w	7/11/2019	2.53
92	51239			Healthy	M	67	99%		2019	right arm	6w	7/31/2019	2.33
93	51093			Healthy	F	77	99%		2019	left arm	6w	6/19/2019	3.08
94	51148			Healthy	M	65	90%		2019	upper arm	6w	8/13/2019	2.21
95	50492			*LRRK2* PD	M	51	91%		2019	left upper arm	6w	7/18/2019	3.55
96	50626			Sporadic PD	M	70	97%	43	2013	unspecified	6w	9/3/2019	2.60

Supplementary Table 1 | Information about the 96 cell lines from 91 donors used in the study. Columns left to right: a 2-digit ID mapping to a cell line from a unique skin biopsy; the biopsy donor; the cross-validation fold for healthy vs. PD prediction; ID for PD individual and matched healthy control; PD status; donor sex; donor age; an ancestry score from genotyping; the Unified Parkinson Disease Rating Scale, a clinical measure of PD severity; skin biopsy collection year; location where biopsy was acquired; cell expansion in 6-well ("6w") or from 12-well to 6-well format ("12w→6w"); date expanded cells were frozen; and doubling time during cell expansion in days, respectively. *: unconfirmed cell line (see Methods).

			Cross-validation							
			set #1	set #2	set #3	set #4	set #5	set #6	set #7	set #8
Batch	Plate layout	Cell lines								
1	1	all 96	test	ignore	ignore	ignore	ignore	train	train	train
	2	all 96	ignore	train	train	train	test	ignore	ignore	ignore
2	1	all 96	ignore	test	ignore	ignore	train	ignore	train	train
	2	all 96	train	ignore	train	train	ignore	test	ignore	ignore
3	1	all 96	ignore	ignore	test	ignore	train	train	ignore	train
	2	all 96	train	train	ignore	train	ignore	ignore	test	ignore
4	1	all 96	ignore	ignore	ignore	test	train	train	train	ignore
	2	all 96	train	train	train	ignore	ignore	ignore	ignore	test

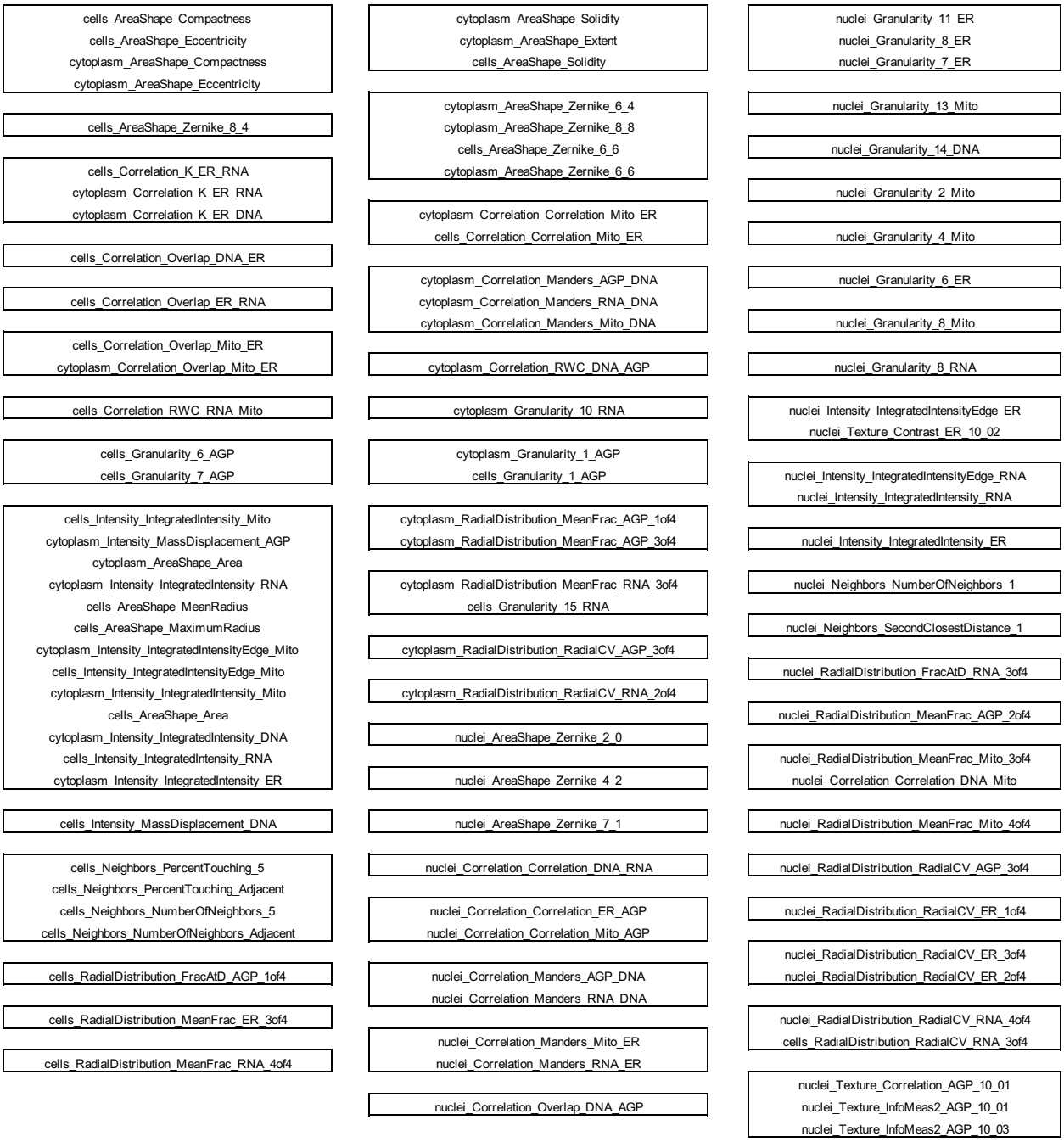
Supplementary Table 2 | Cross-validation strategy for 96-way cell line classification. For each of 8 cross-validation sets, both batch and plate layout were held out in the test set.

			Cross-validation							
			set #1	set #2	set #3	set #4	set #5	set #6	set #7	set #8
Batch	Plate layout	Cell lines								
1	1	5 held-out biopsies	test	ignore	ignore	ignore	ignore	ignore	ignore	ignore
		remaining 91 lines	ignore	ignore	ignore	ignore	ignore	train	train	train
	2	5 held-out biopsies	ignore	ignore	ignore	ignore	test	ignore	ignore	ignore
		remaining 91 lines	ignore	train	train	train	ignore	ignore	ignore	ignore
2	1	5 held-out biopsies	ignore	test	ignore	ignore	ignore	ignore	ignore	ignore
		remaining 91 lines	ignore	ignore	ignore	ignore	train	ignore	train	train
	2	5 held-out biopsies	ignore	ignore	ignore	ignore	ignore	test	ignore	ignore
		remaining 91 lines	train	ignore	train	train	ignore	ignore	ignore	ignore
3	1	5 held-out biopsies	ignore	ignore	test	ignore	ignore	ignore	ignore	ignore
		remaining 91 lines	ignore	ignore	ignore	ignore	train	train	ignore	train
	2	5 held-out biopsies	ignore	ignore	ignore	ignore	ignore	ignore	test	ignore
		remaining 91 lines	train	train	ignore	train	ignore	ignore	ignore	ignore
4	1	5 held-out biopsies	ignore	ignore	ignore	test	ignore	ignore	ignore	ignore
		remaining 91 lines	ignore	ignore	ignore	ignore	train	train	train	ignore
	2	5 held-out biopsies	ignore	ignore	ignore	ignore	ignore	ignore	ignore	test
		remaining 91 lines	train	train	train	ignore	ignore	ignore	ignore	ignore

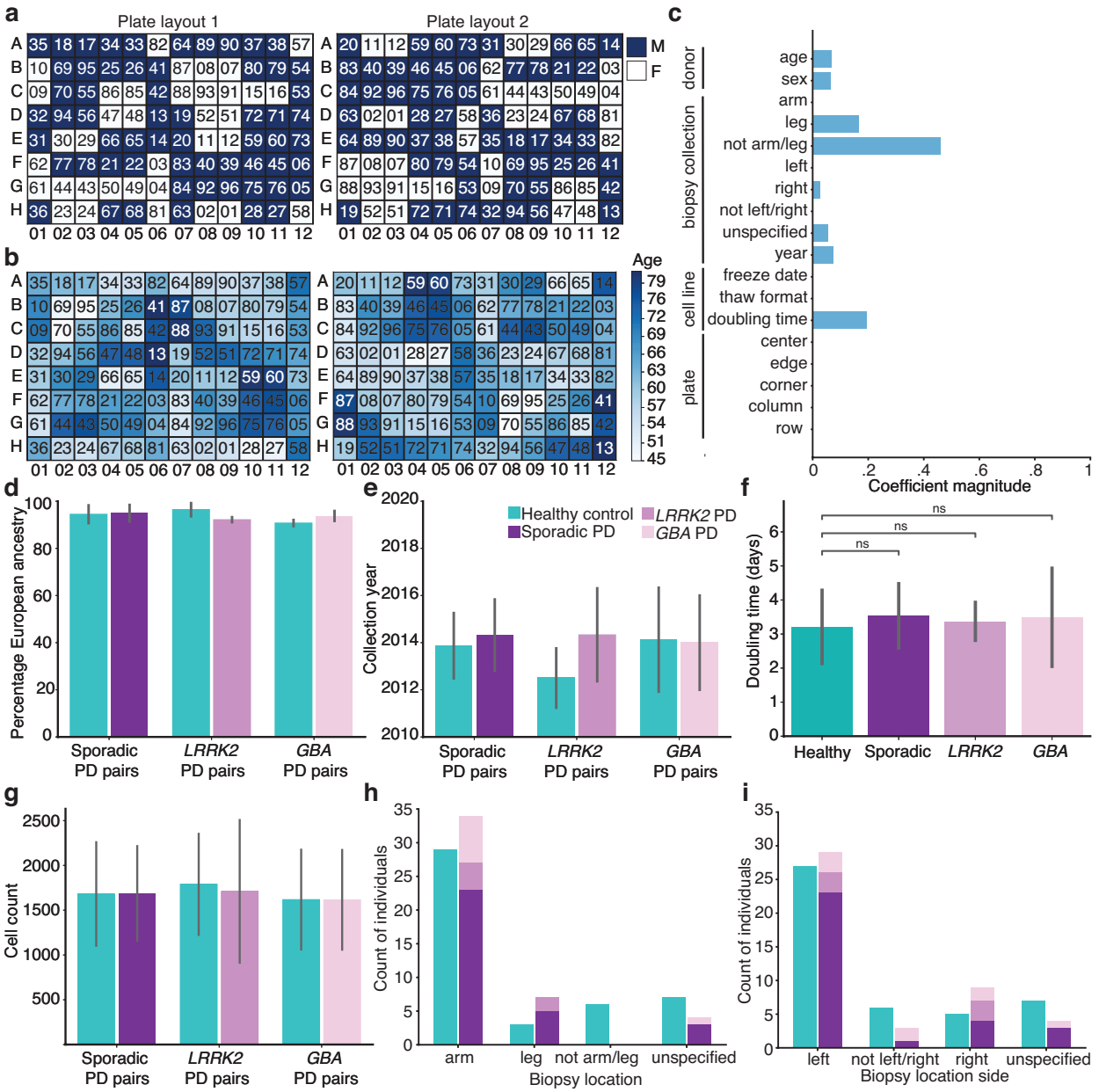
Supplementary Table 3 | Cross-validation strategy for 91-way biopsy donor classification. For each of 8 cross-validation sets, the test set consisted of cell lines from one of the two biopsies from the 5 individuals who donated two biopsies, while the train set consisted of cell lines from the complementary set of biopsies from these 5 individuals and the remaining 86 individuals who donated only a single biopsy. To avoid plate position biases as potential confounds, plate layout was also held out, and to assess model generalization to a test biopsy acquired in a new batch, batch was also held out. These 8 cross-validation sets were conducted twice, once holding out in the test sets the earlier set of skin biopsies from the 5 individuals who donated two biopsies (cell lines 08, 39, 51, 55, 70), and again holding out the later set (cell lines 91, 92, 93, 94, 95).

					Cross-validation				
					set #0	set #1	set #2	set #3	set #4
Cell line	Disease state	Pair ID	Batch	Plate layout					
25	Healthy	12	all 4	both	test	train	train	train	train
26	Sporadic PD	12	all 4	both	test	train	train	train	train
33	Healthy	16	all 4	both	test	train	train	train	train
34	Sporadic PD	16	all 4	both	test	train	train	train	train
39	Healthy	19	all 4	both	test	train	train	train	train
40	Sporadic PD	19	all 4	both	test	train	train	train	train
43	Healthy	21	all 4	both	test	train	train	train	train
44	Sporadic PD	21	all 4	both	test	train	train	train	train
51	Healthy	25	all 4	both	test	train	train	train	train
52	Sporadic PD	25	all 4	both	test	train	train	train	train
71	Healthy	35	all 4	both	test	train	train	train	train
72	Sporadic PD	35	all 4	both	test	train	train	train	train
81	Healthy	40	all 4	both	test	train	train	train	train
82	*LRRK2* PD	40	all 4	both	test	train	train	train	train
19	Healthy	9	all 4	both	test	train	train	train	train
20	Sporadic PD	9	all 4	both	test	train	train	train	train
1	Healthy	0	all 4	both	train	test	train	train	train
2	*LRRK2* PD	0	all 4	both	train	test	train	train	train
21	Healthy	10	all 4	both	train	test	train	train	train
22	Sporadic PD	10	all 4	both	train	test	train	train	train
35	Healthy	17	all 4	both	train	test	train	train	train
36	Sporadic PD	17	all 4	both	train	test	train	train	train
41	Healthy	20	all 4	both	train	test	train	train	train
42	Sporadic PD	20	all 4	both	train	test	train	train	train
61	Healthy	30	all 4	both	train	test	train	train	train
62	Sporadic PD	30	all 4	both	train	test	train	train	train
69	Healthy	34	all 4	both	train	test	train	train	train
70	*LRRK2* PD	34	all 4	both	train	test	train	train	train
9	Healthy	4	all 4	both	train	test	train	train	train
10	Sporadic PD	4	all 4	both	train	test	train	train	train
3	Healthy	1	all 4	both	train	train	test	train	train
4	Sporadic PD	1	all 4	both	train	train	test	train	train
29	Healthy	14	all 4	both	train	train	test	train	train
30	Sporadic PD	14	all 4	both	train	train	test	train	train
31	Healthy	15	all 4	both	train	train	test	train	train
32	Sporadic PD	15	all 4	both	train	train	test	train	train
45	Healthy	22	all 4	both	train	train	test	train	train
46	*LRRK2* PD	22	all 4	both	train	train	test	train	train
49	Healthy	24	all 4	both	train	train	test	train	train
50	Sporadic PD	24	all 4	both	train	train	test	train	train
67	Healthy	33	all 4	both	train	train	test	train	train
68	Sporadic PD	33	all 4	both	train	train	test	train	train
89	Healthy	44	all 4	both	train	train	test	train	train
90	Sporadic PD	44	all 4	both	train	train	test	train	train
37	Healthy	18	all 4	both	train	train	train	test	train
38	*LRRK2* PD	18	all 4	both	train	train	train	test	train
59	Healthy	29	all 4	both	train	train	train	test	train
60	Sporadic PD	29	all 4	both	train	train	train	test	train
63	Healthy	31	all 4	both	train	train	train	test	train
64	Sporadic PD	31	all 4	both	train	train	train	test	train
85	Healthy	42	all 4	both	train	train	train	test	train
86	Sporadic PD	42	all 4	both	train	train	train	test	train
87	Healthy	43	all 4	both	train	train	train	test	train
88	Sporadic PD	43	all 4	both	train	train	train	test	train
11	Healthy	5	all 4	both	train	train	train	test	train
12	Sporadic PD	5	all 4	both	train	train	train	test	train
17	Healthy	8	all 4	both	train	train	train	test	train
18	Sporadic PD	8	all 4	both	train	train	train	test	train
5	Healthy	2	all 4	both	train	train	train	train	test
6	Sporadic PD	2	all 4	both	train	train	train	train	test
55	Healthy	27	all 4	both	train	train	train	train	test
56	Sporadic PD	27	all 4	both	train	train	train	train	test
57*	Healthy*	28	all 4	both	train	train	train	train	test
58	Sporadic PD	28	all 4	both	train	train	train	train	test
7	Healthy	3	all 4	both	train	train	train	train	test
8	Sporadic PD	3	all 4	both	train	train	train	train	test
65	Healthy	32	all 4	both	train	train	train	train	test
66	*LRRK2* PD	32	all 4	both	train	train	train	train	test
83	Healthy	41	all 4	both	train	train	train	train	test
84	Sporadic PD	41	all 4	both	train	train	train	train	test
13	Healthy	6	all 4	both	train	train	train	train	test
14	Sporadic PD	6	all 4	both	train	train	train	train	test
15	Healthy	7	all 4	both	train	train	train	train	test
16	Sporadic PD	7	all 4	both	train	train	train	train	test

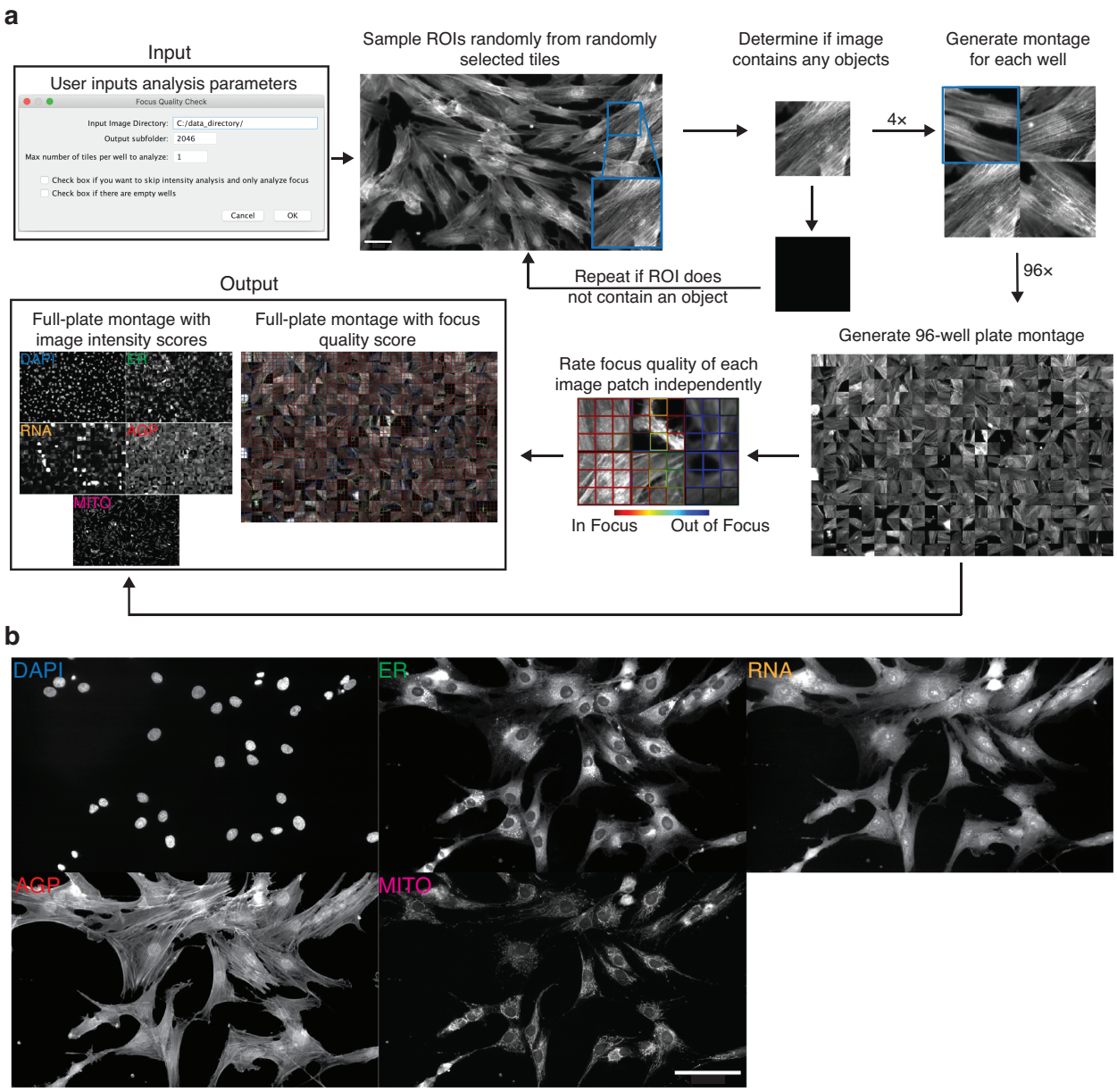
Supplementary Table 4 | 5-fold Cross-validation strategy for healthy vs. PD classification. A subset of 74 cell lines from 74 individuals (6 *LRRK2* PD and paired controls, and 31 sporadic PD and paired controls was divided into 5 cross-validation folds. For each of 5 cross-validation sets, one fold of cell lines was held out in the test set. *: unconfirmed cell line (see Methods).



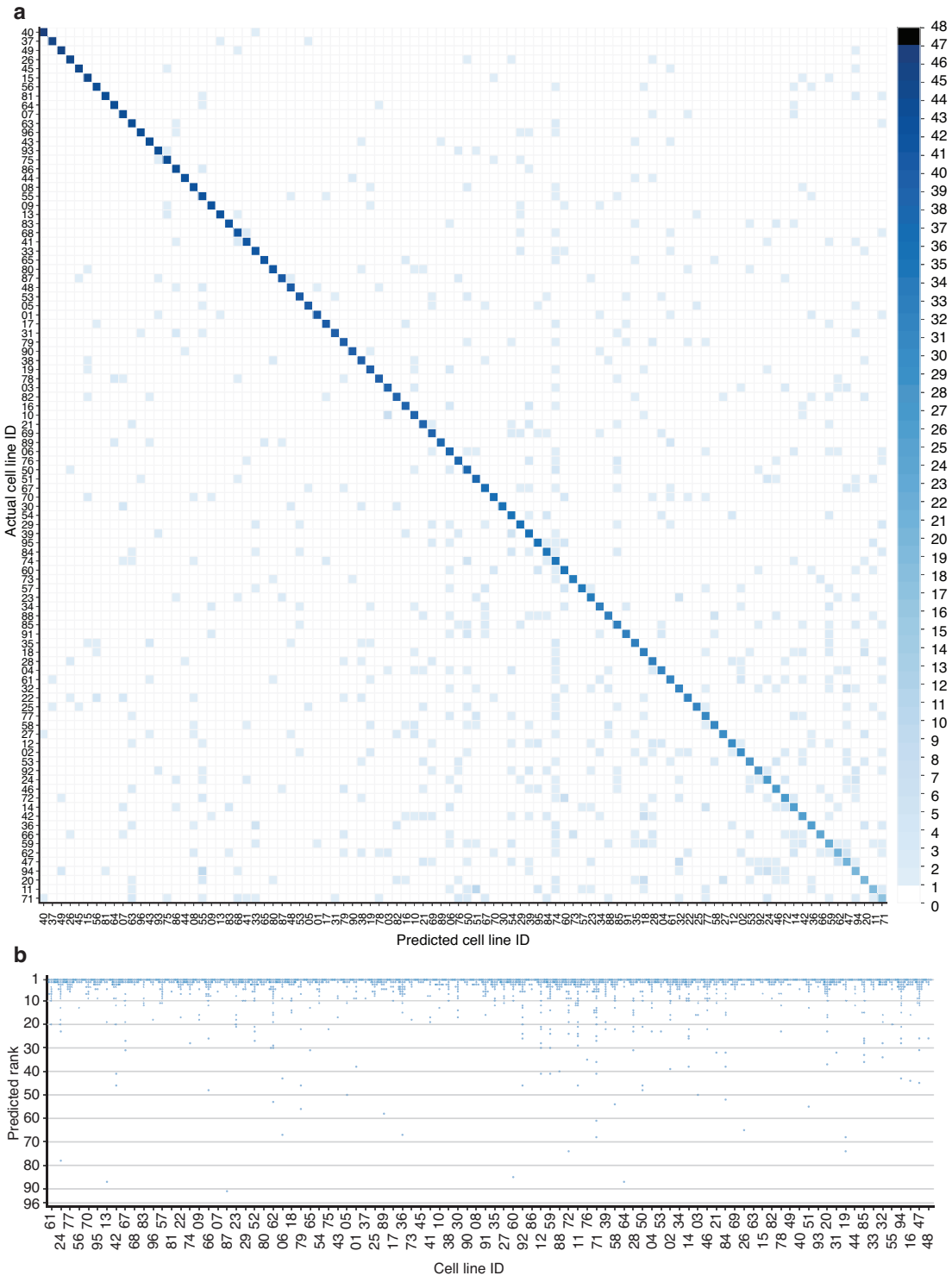
Supplementary Table 5 | Most common important CellProfiler features grouped based on correlation. The top 100 most important CellProfiler features from **Fig. 6a**, clustered into 55 groups based on Pearson correlation.



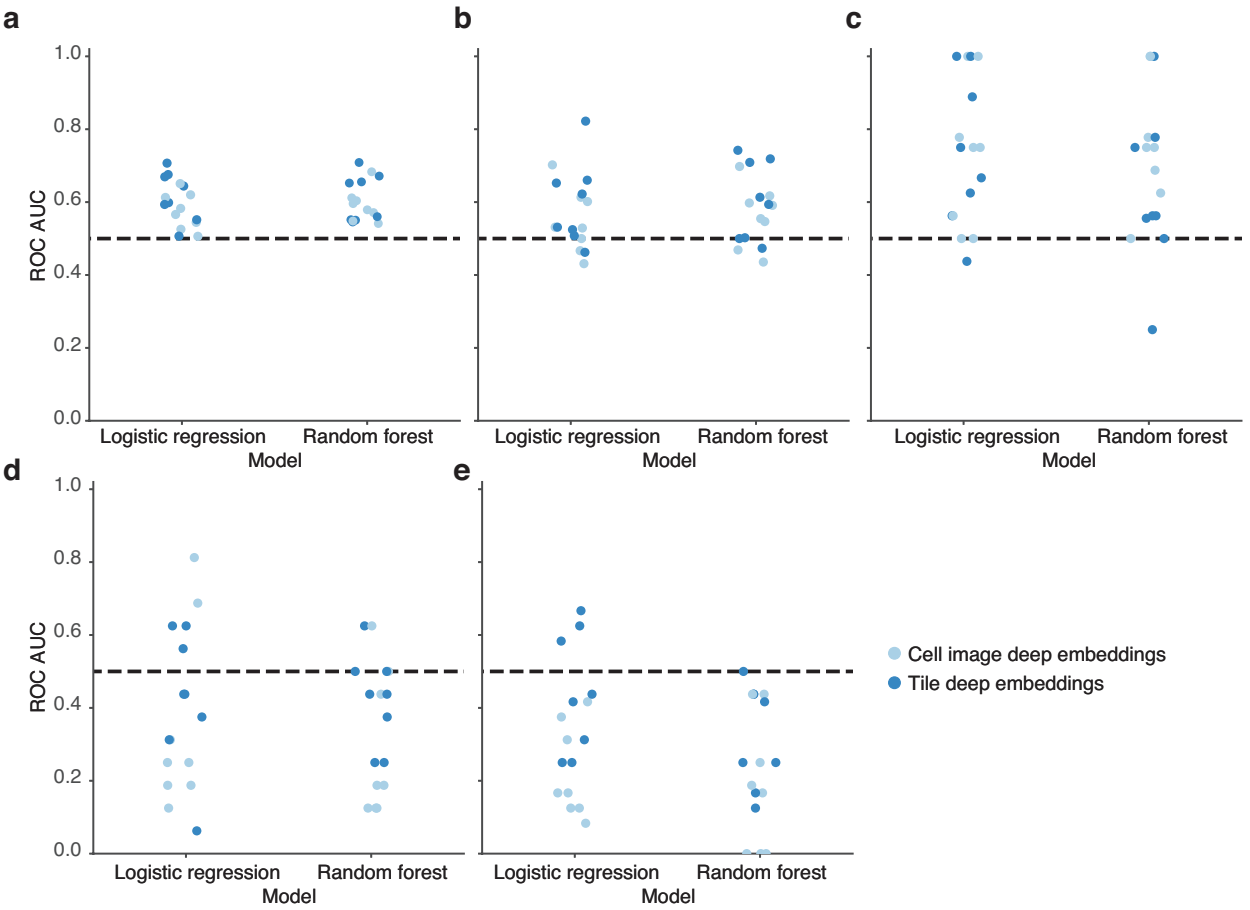
Supplementary Fig. 1 | Experiment design details for high-content screening. Various donor demographics including (**a**) sex (male (M), female (F)) and (**b**) age for the two 96-well plate layouts, where each well contains cells from the cell line denoted by the two-digit label. **c**, Lasso variable selection for healthy vs. PD on donor, biopsy, cell line, and plate covariates reveals no significant biases. Distributions of additional cell line covariates including (**d**) percentage European ancestry from genotyping analysis, (**e**) biopsy collection year, (**f**) cell doubling times (Mann–Whitney U = 57.0, p = 1.0 × 10^{-2} for sporadic, U = 118.0, p = 6.4 × 10^{-1} for *LRRK2 PD*, and U = 193.5, p = 1.00 for *GBA PD* vs. healthy, respectively, ns: p > 5.0 × 10^{-2}), (**g**) well-level cell count, and biopsy location, (**h**) arm or leg and (**i**) left or right. Error bars denote standard deviation.



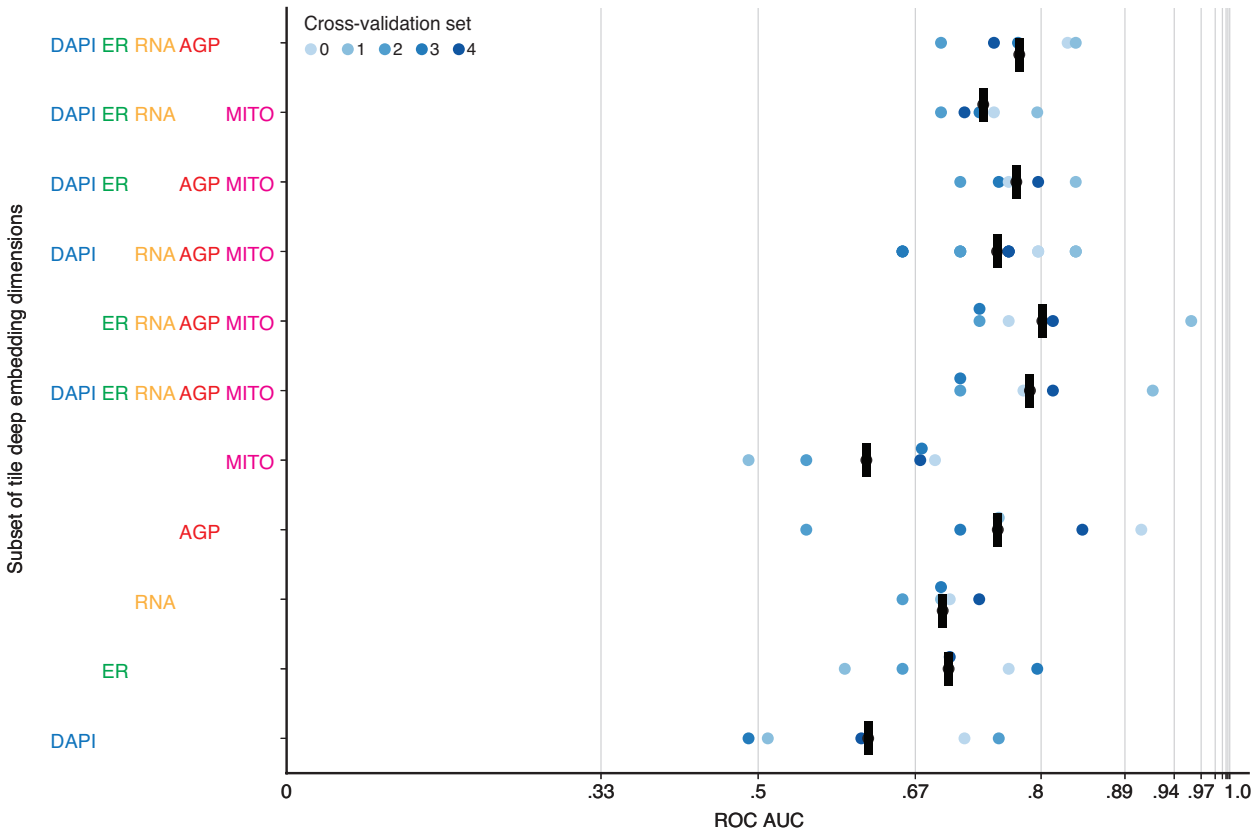
Supplementary Fig. 2 | Overview of near real-time image quality analysis and sample Cell Painting images of primary human fibroblasts. A Fiji (an ImageJ distribution) macro assesses the quality and consistency of the images sampled from a full 96-well plate. **a**, Four random regions of interest (ROI) are cropped from images in each channel and in each well, and 96-well montages are constructed for viewing. A measurement of mean image intensity across the plate is reported for each plate montage. Next, the montage corresponding to the user-designated focus channel is inputted to a microscope image focus classifier which calculates a focus quality score for each image patch. For visualization, a color-coded overlay on top of the montage highlights regions that are in focus (red) or out of focus (blue). Scale bar: 50 μm. **b**, Sample images of one tile from the 5 Cell Painting channels. Scale bar: 100 μm.



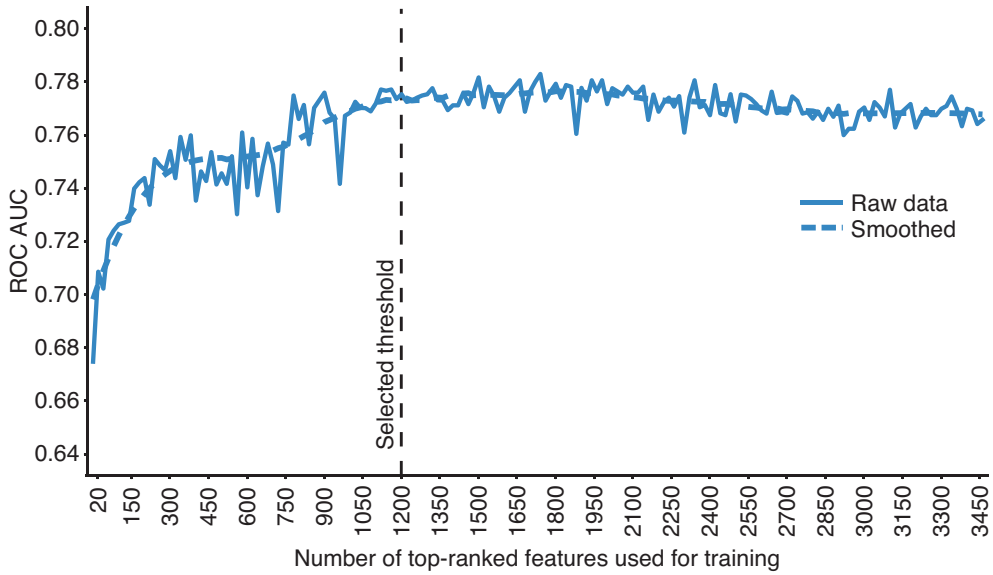
Supplementary Fig. 3 | Identification of individual cell lines in held-out batches and plate layouts at the well-level. a, Confusion matrix, sorted by the diagonal, showing the test set well-level predicted and actual cell lines for each of 6 wells in each of 8 held-out batch and held-out plate layouts for the model in **Fig. 3c**. **b**, Test set well-level predicted rank, among 96 of the 6 wells in each of 8 held-out batch and held-out plate layouts for the model in **Fig. 3c**.



Supplementary Fig. 4 | Preliminary evaluation of PD classification performance. Test set cell line–level PD classification for (**a**) all PD (*n* = 45 participants) and matched controls (n = 45 participants), (**b**) sporadic PD (*n* = 31) and matched controls (*n* = 31 participants), (**c**) *LRRK2* PD (*n* = 6 participants) and matched controls (*n* = 6 participants), (**d**) *GBA* PD (*n* = 8 participants) and matched controls (*n* = 8 participants), and (**e**) *GBA* PD (*n* = 7 participants) and matched controls (*n* = 8 participants), excluding the unconfirmed GBA line (see Methods). In each case, for cross-validation, matched cell line pairs were randomly divided into a train half and a test half 8 times. Dashed line denotes chance performance.



Supplementary Fig. 5 | Impact of individual Cell Painting channels on PD classification. The same logistic regression model with tile deep embeddings from **Fig. 5b** evaluated with a subset of the deep embedding dimensions corresponding to a subset of the 5 channels. Black bars denote the mean across all cross-validation sets. Grid line spacing denotes a doubling of the odds of correctly ranking a random healthy control and PD cell line. Dashed line denotes chance performance.



Supplementary Fig. 6 | Estimating threshold for number of top-ranked CellProfiler features required for PD classification. Performance of the random forest classifier as a function of number of top-ranked features used for training, evaluated in increments of 20 features. The dashed line represents the threshold selected for subsequent analyses.