Electronic Supplementary Information

A Density-Watershed Algorithm (DWA) Method for Robust, Accurate and Automatic Classification of Dual-Fluorescence and Four-Cluster Droplet Digital PCR Data

Xiurui Zhu[1], Shisheng Su[1], Mingzhu Fu[1], Zhiyong Peng[2], Dong Wang[2], Xiao Rui[2], Fang Wang[1], Xiaobin Liu[2], Baoxia Liu[2], Lingxiang Zhu[1,3], Wenjun Yang[1,2], Na Gao[2], Guoliang Huang[1,4], Gaoshan Jing[5,6,*], Yong Guo[1,7,*]

Author affiliations:

[1]Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China

[2]TargetingOne Corporation, Beijing, China.

[3]National Research Institute for Family Planning, Beijing, China.

[4]National Engineering Research Center for Beijing Biochip Technology, Beijing, China.

[5]Department of Precision Instrument, School of Mechanical Engineering, Tsinghua University, Beijing, China

[6]State Key Laboratory of Precision Measurement Technology and Instruments, Beijing, China.

[7]Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Beijing, China.

** Corresponding authors:*

*Correspondence and requests for materials should be addressed to

Dr. Gaoshan Jing

Email: gaoshanjing@mail.tsinghua.edu.cn

or

Dr. Yong Guo

Email: yongguo@tsinghua.edu.cn

Contents

Supplementary Figures S1–7: *EGFR* L858R Classification Results (Plasmid Samples)

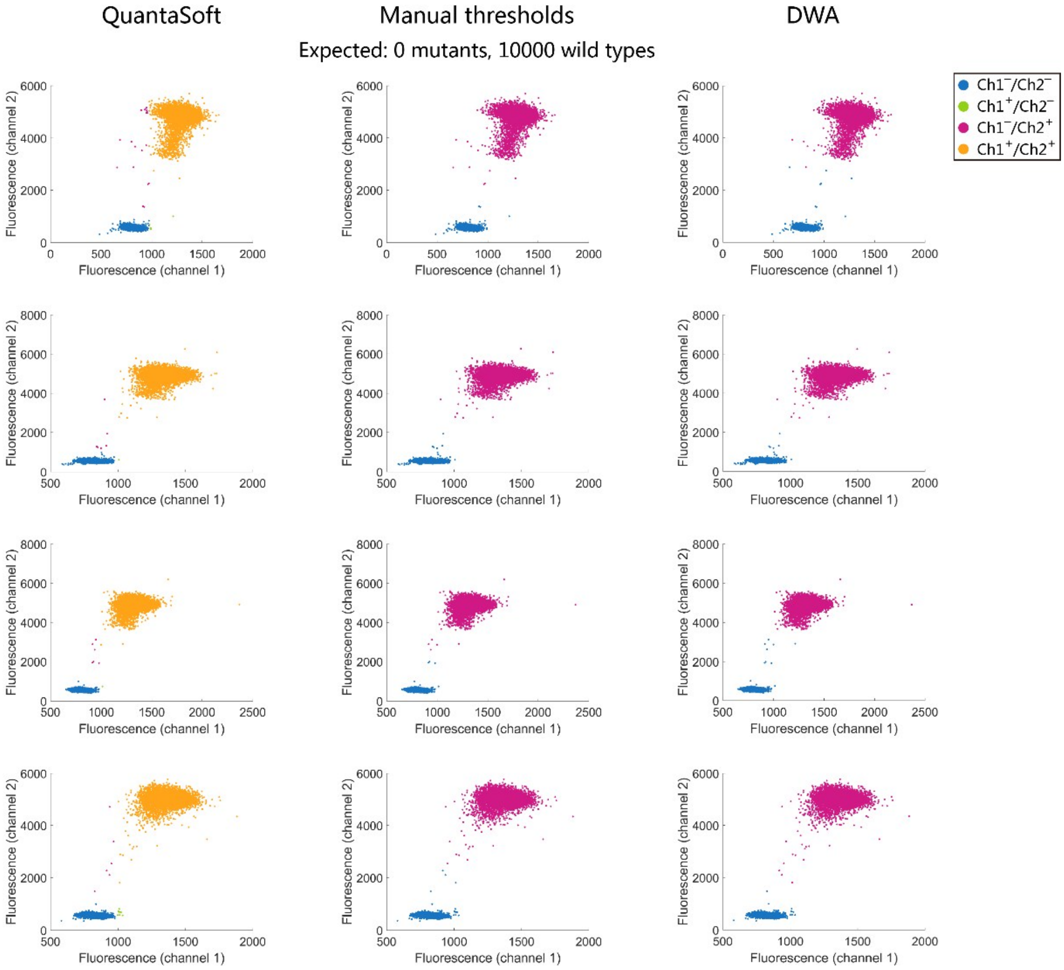


Figure S1. Comparison between classification results by QuantaSoft's automatic mode, manual thresholds and the DWA method. Each *EGFR* L858R assay contains 0 copies of mutants (Ch1) and about 10000 copies of wild types (Ch2). Each row is a replicate.

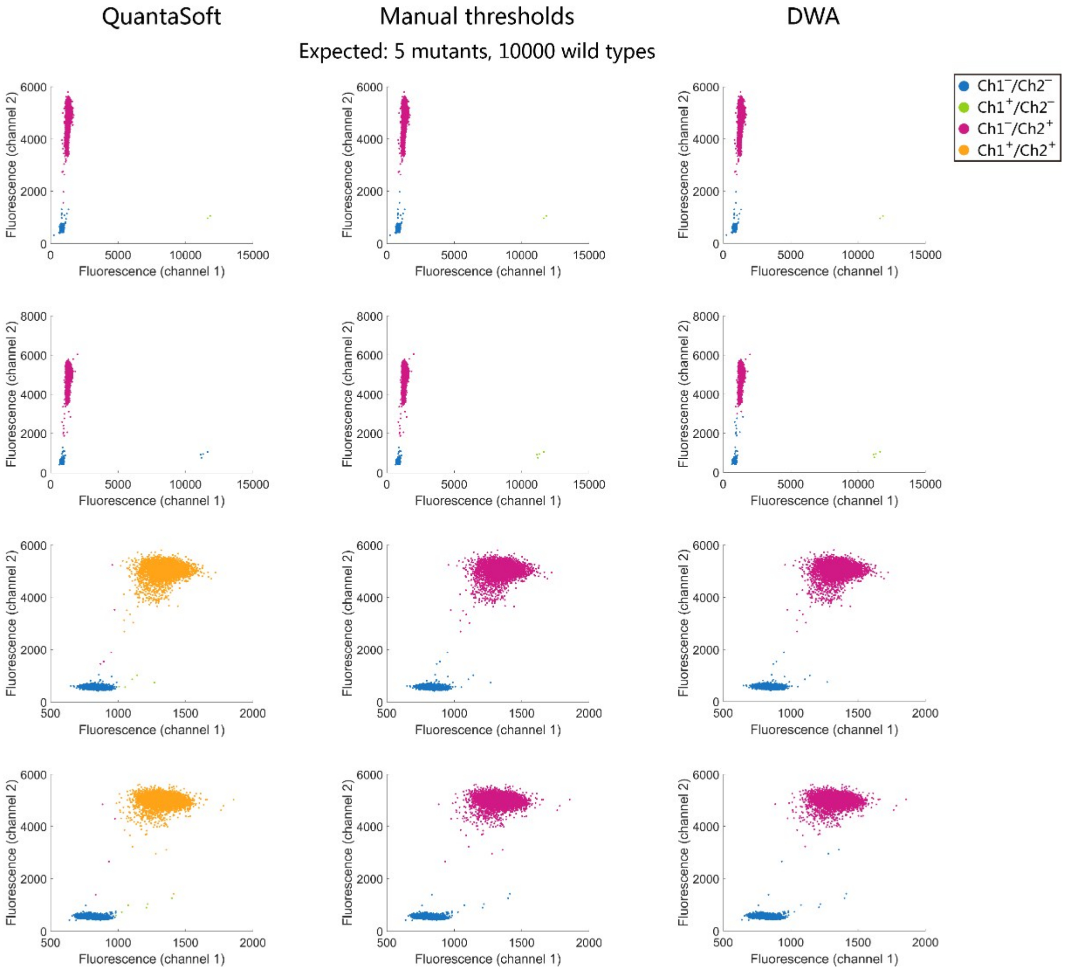


Figure S2. Comparison between classification results by QuantaSoft's automatic mode, manual thresholds and the DWA method. Each *EGFR* L858R assay contains about 5 copies of mutants (Ch1) and about 10000 copies of wild types (Ch2). Each row is a replicate.

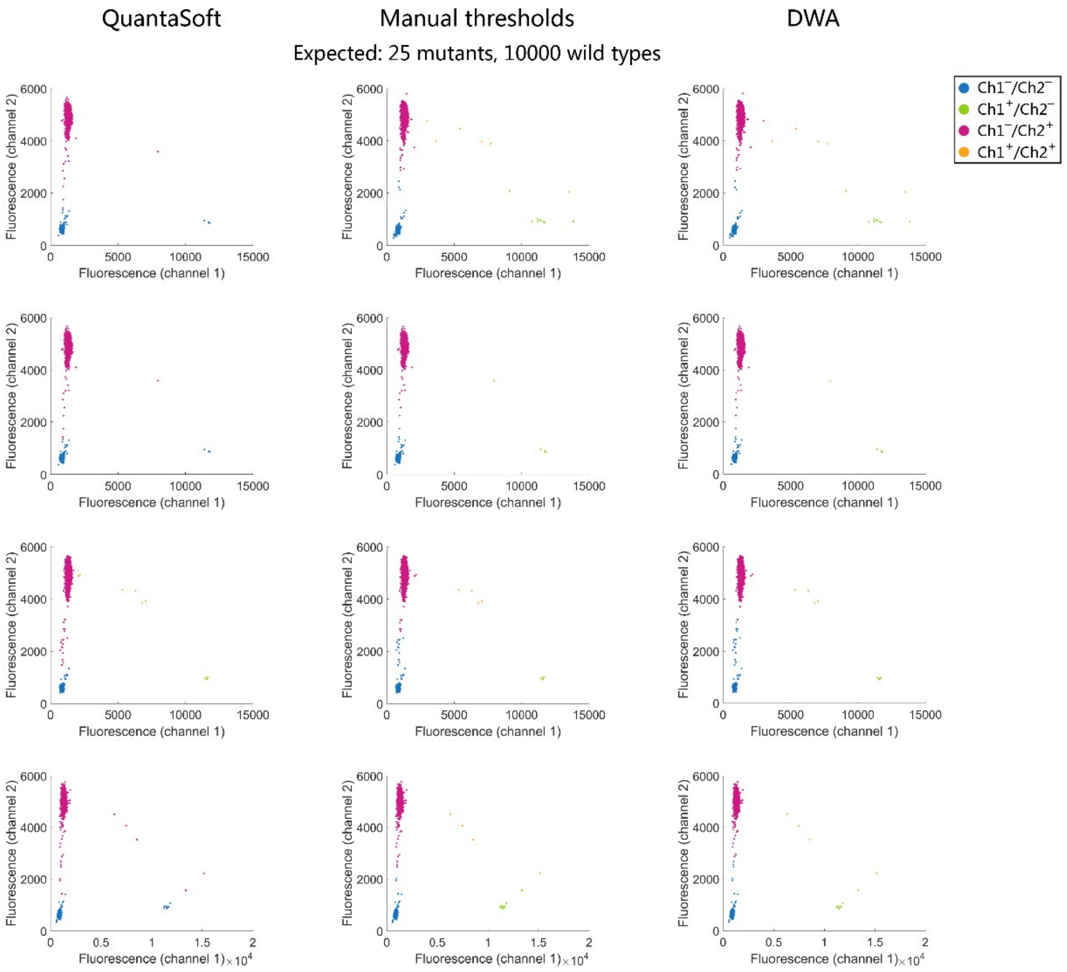


Figure S3. Comparison between classification results by QuantaSoft's automatic mode, manual thresholds and the DWA method. Each *EGFR* L858R assay contains about 25 copies of mutants (Ch1) and about 10000 copies of wild types (Ch2). Each row is a replicate.

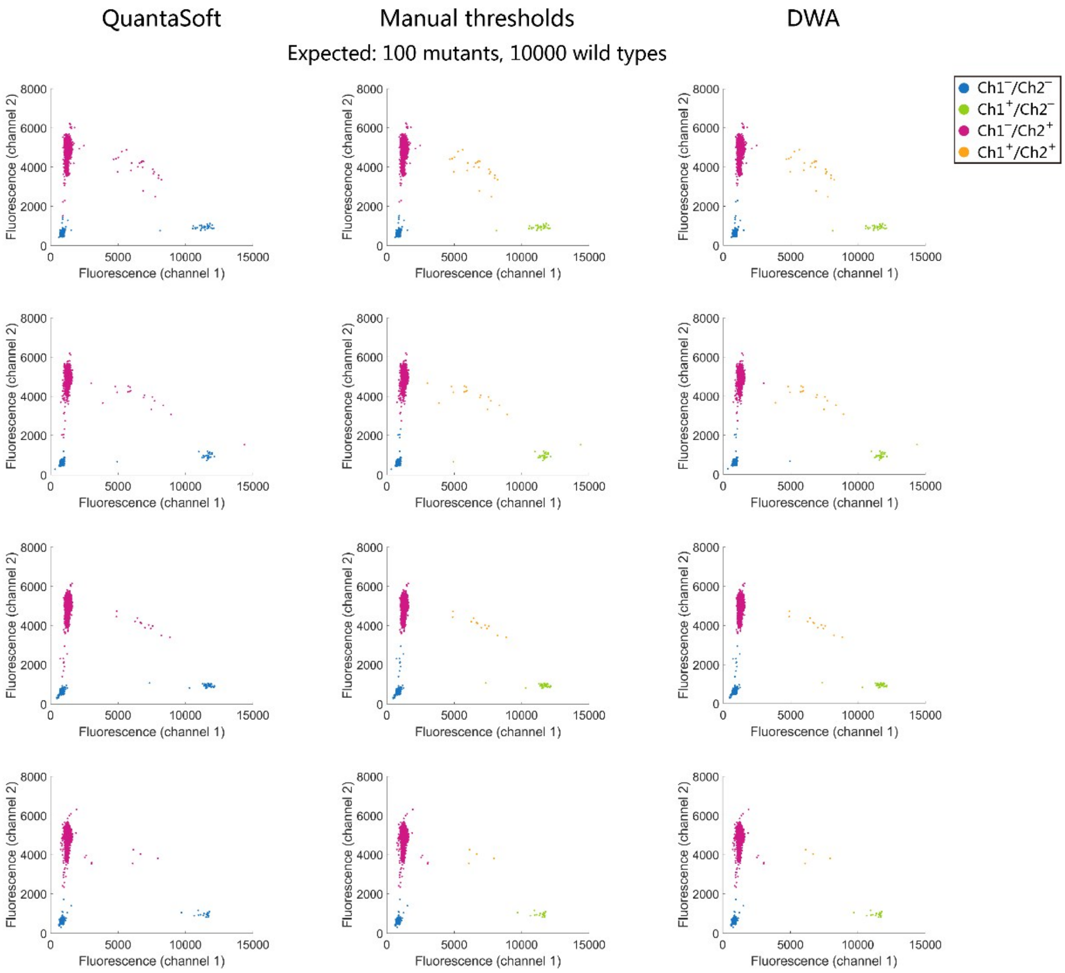


Figure S4. Comparison between classification results by QuantaSoft's automatic mode, manual thresholds and the DWA method. Each *EGFR* L858R assay contains about 100 copies of mutants (Ch1) and about 10000 copies of wild types (Ch2). Each row is a replicate.

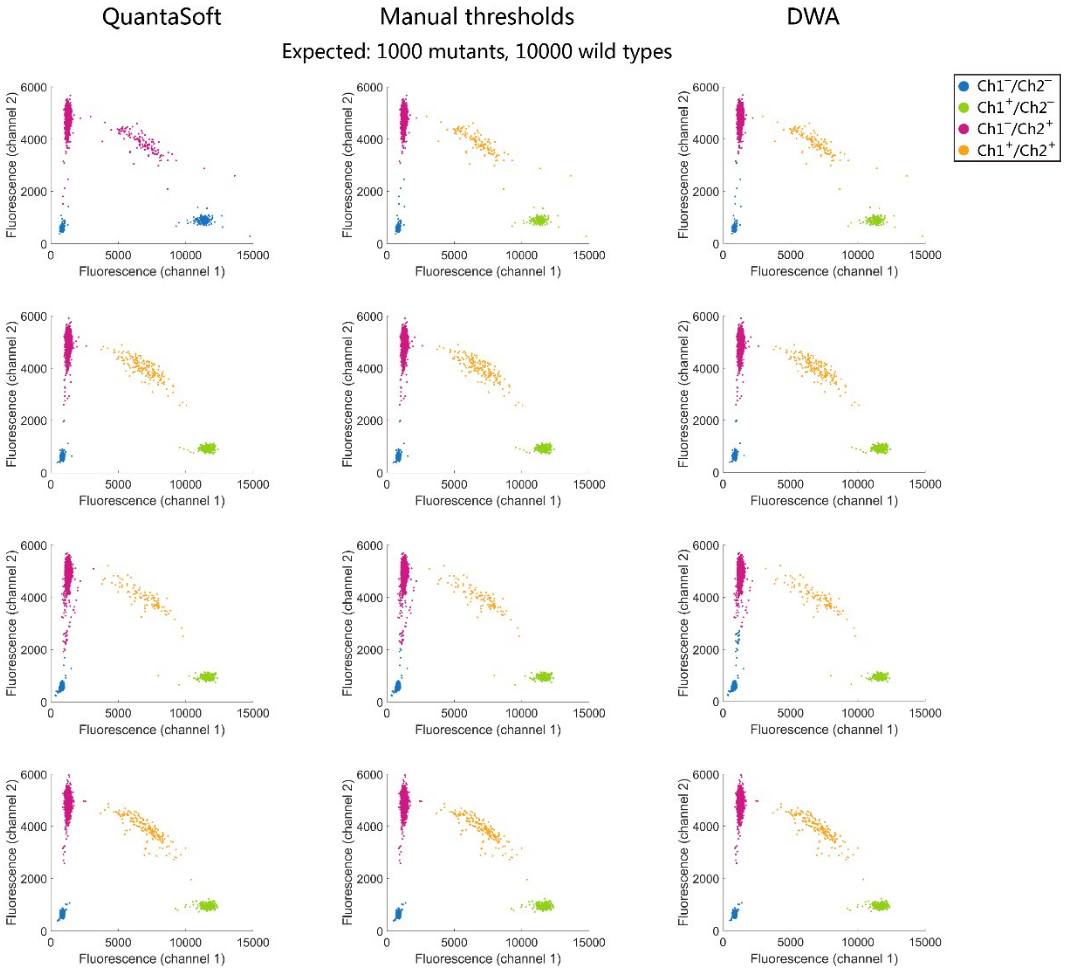


Figure S5. Comparison between classification results by QuantaSoft's automatic mode, manual thresholds and the DWA method. Each *EGFR* L858R assay contains about 1000 copies of mutants (Ch1) and about 10000 copies of wild types (Ch2). Each row is a replicate.

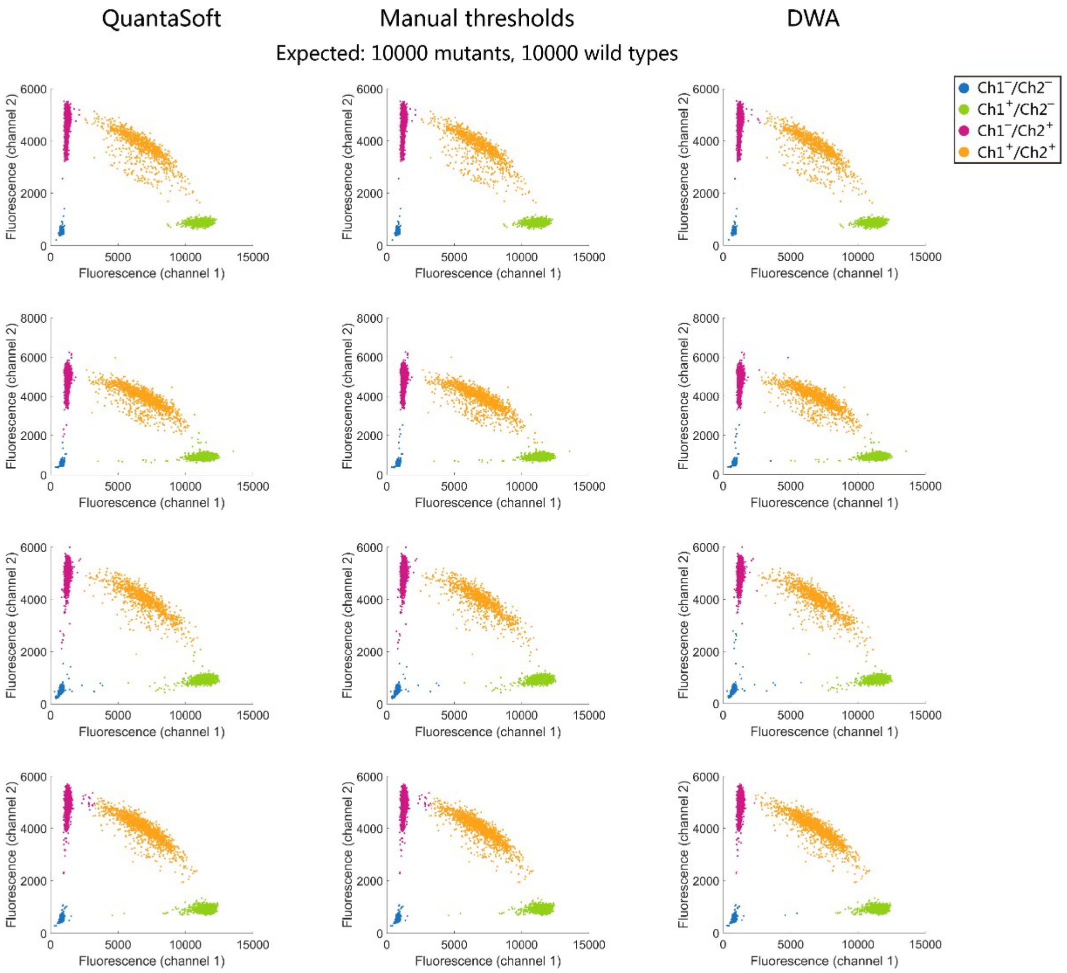


Figure S6. Comparison between classification results by QuantaSoft's automatic mode, manual thresholds and the DWA method. Each *EGFR* L858R assay contains about 10000 copies of mutants (Ch1) and about 10000 copies of wild types (Ch2). Each row is a replicate.

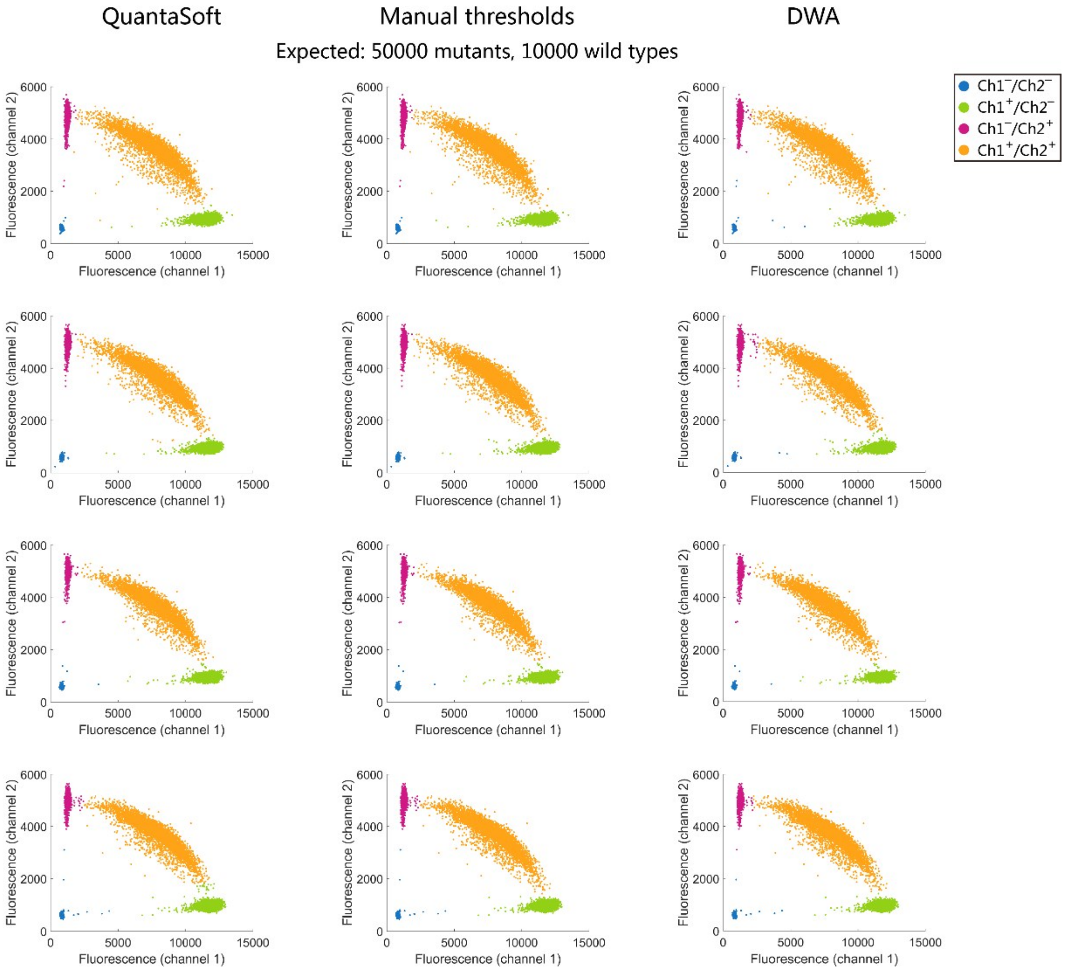


Figure S7. Comparison between classification results by QuantaSoft's automatic mode, manual thresholds and the DWA method. Each *EGFR* L858R assay contains about 50000 copies of mutants (Ch1) and about 10000 copies of wild types (Ch2). Each row is a replicate.

Supplementary Figures S8–14: *EGFR* T790M Classification Results (Plasmid Samples)

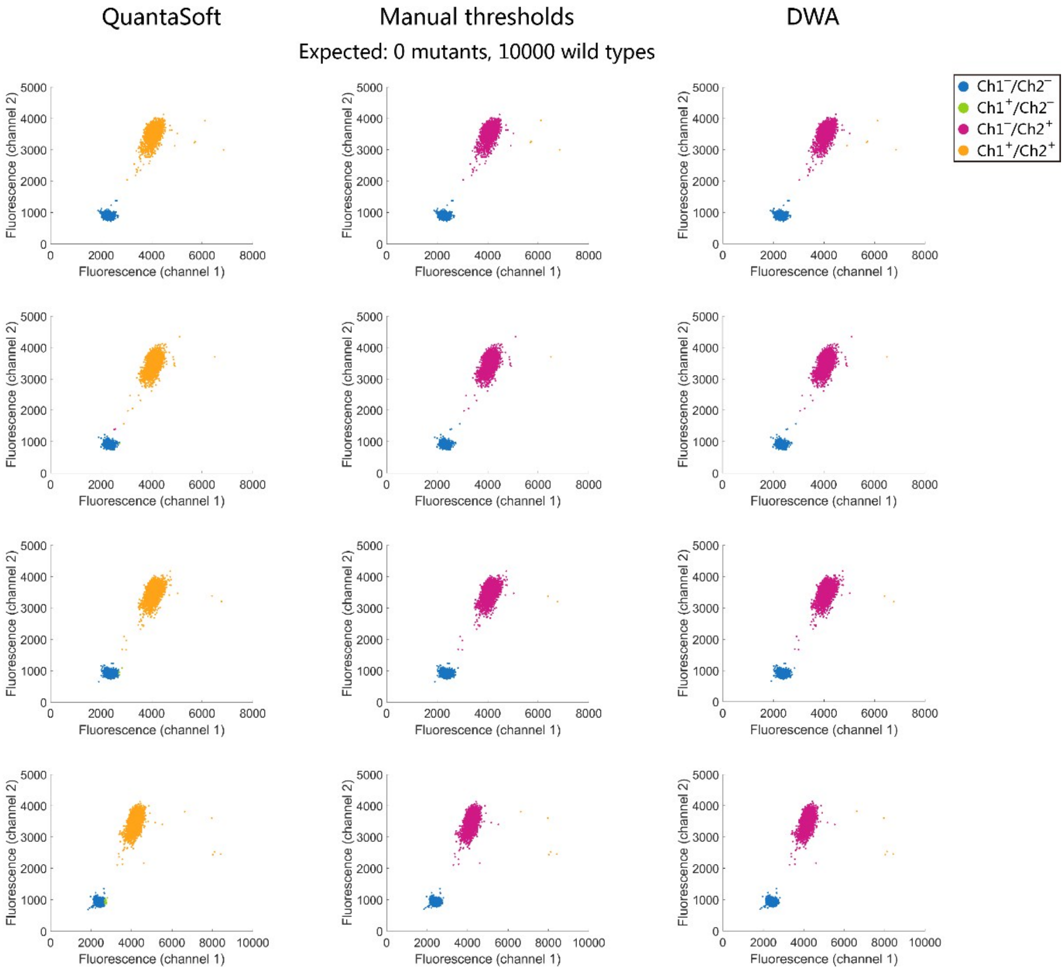


Figure S8. Comparison between classification results by QuantaSoft's automatic mode, manual thresholds and the DWA method. Each *EGFR* T790M assay contains 0 copies of mutants (Ch1) and about 10000 copies of wild types (Ch2). Each row is a replicate.

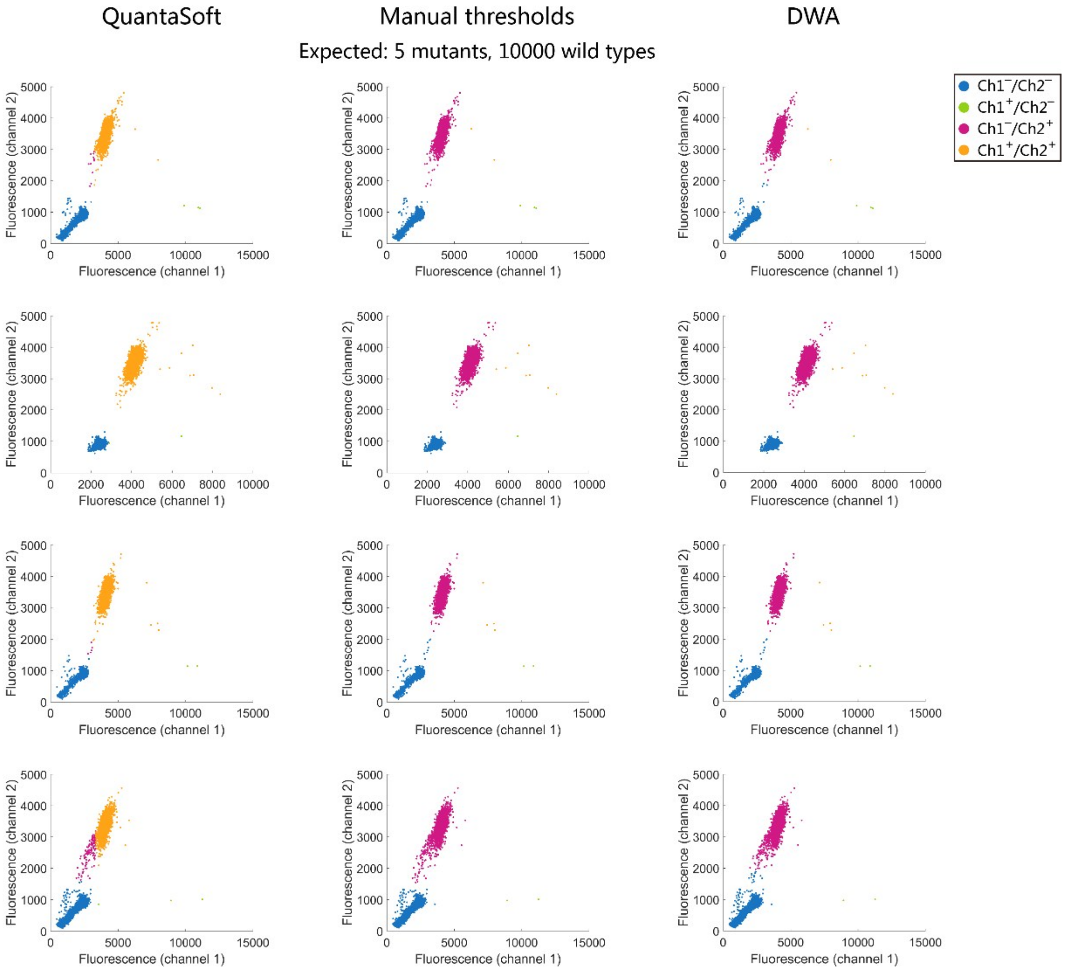


Figure S9. Comparison between classification results by QuantaSoft's automatic mode, manual thresholds and the DWA method. Each *EGFR* T790M assay contains about 5 copies of mutants (Ch1) and about 10000 copies of wild types (Ch2). Each row is a replicate.

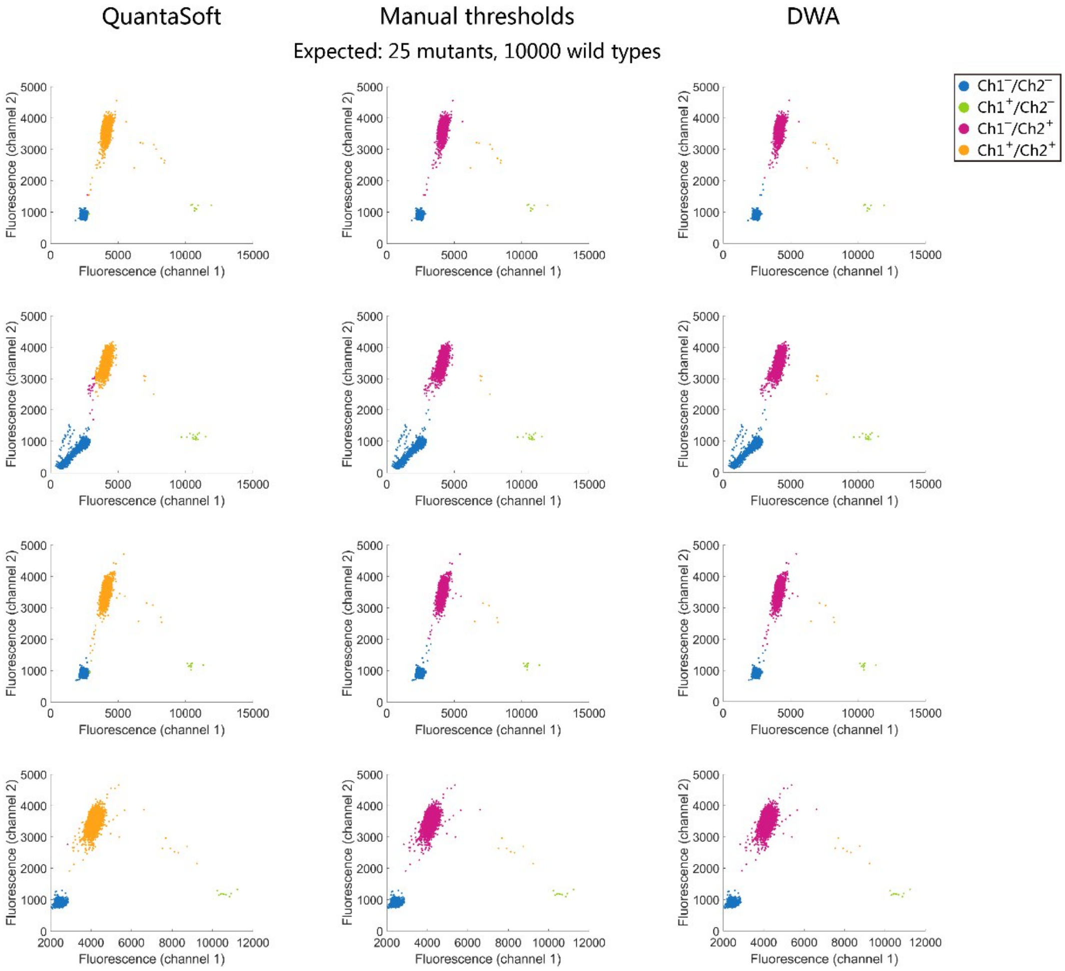


Figure S10. Comparison between classification results by QuantaSoft’s automatic mode, manual thresholds and the DWA method. Each *EGFR* T790M assay contains about 25 copies of mutants (Ch1) and about 10000 copies of wild types (Ch2). Each row is a replicate.

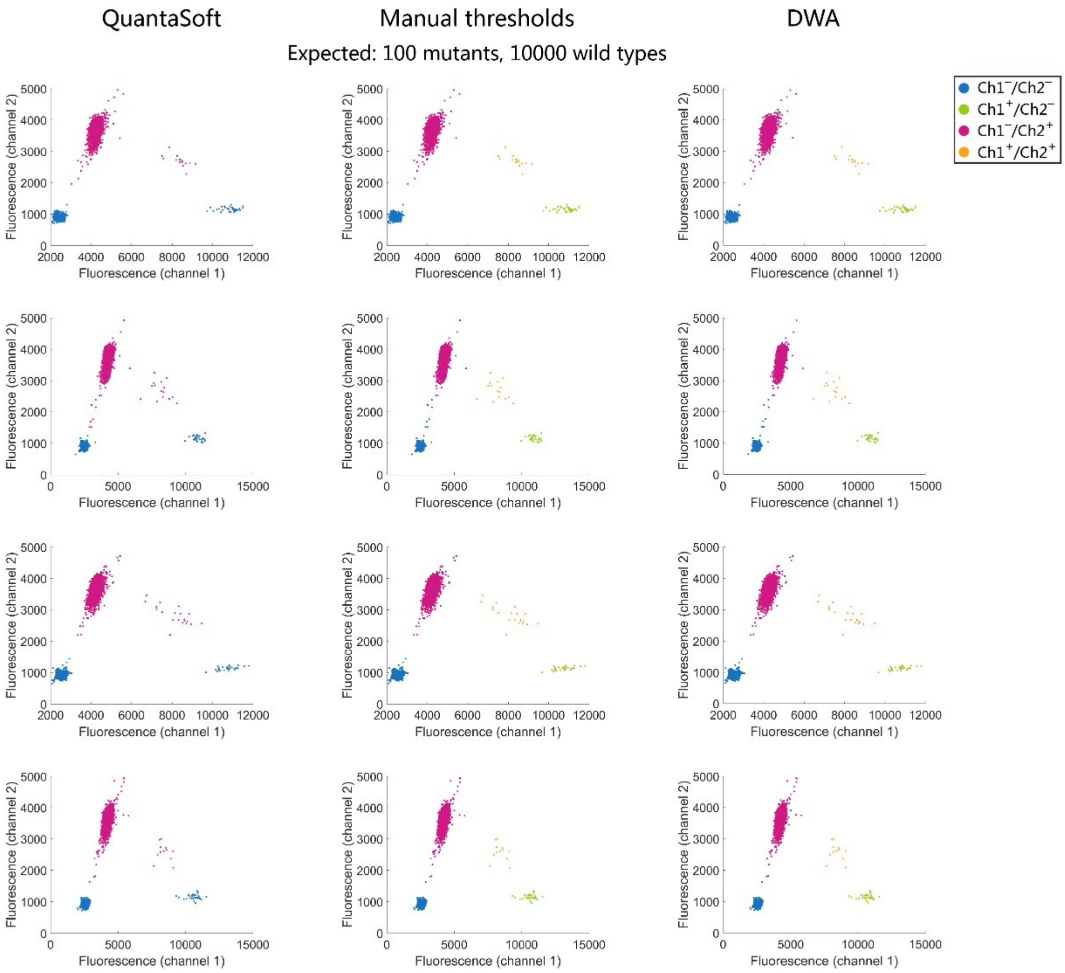


Figure S11. Comparison between classification results by QuantaSoft's automatic mode, manual thresholds and the DWA method. Each *EGFR* T790M assay contains about 100 copies of mutants (Ch1) and about 10000 copies of wild types (Ch2). Each row is a replicate.

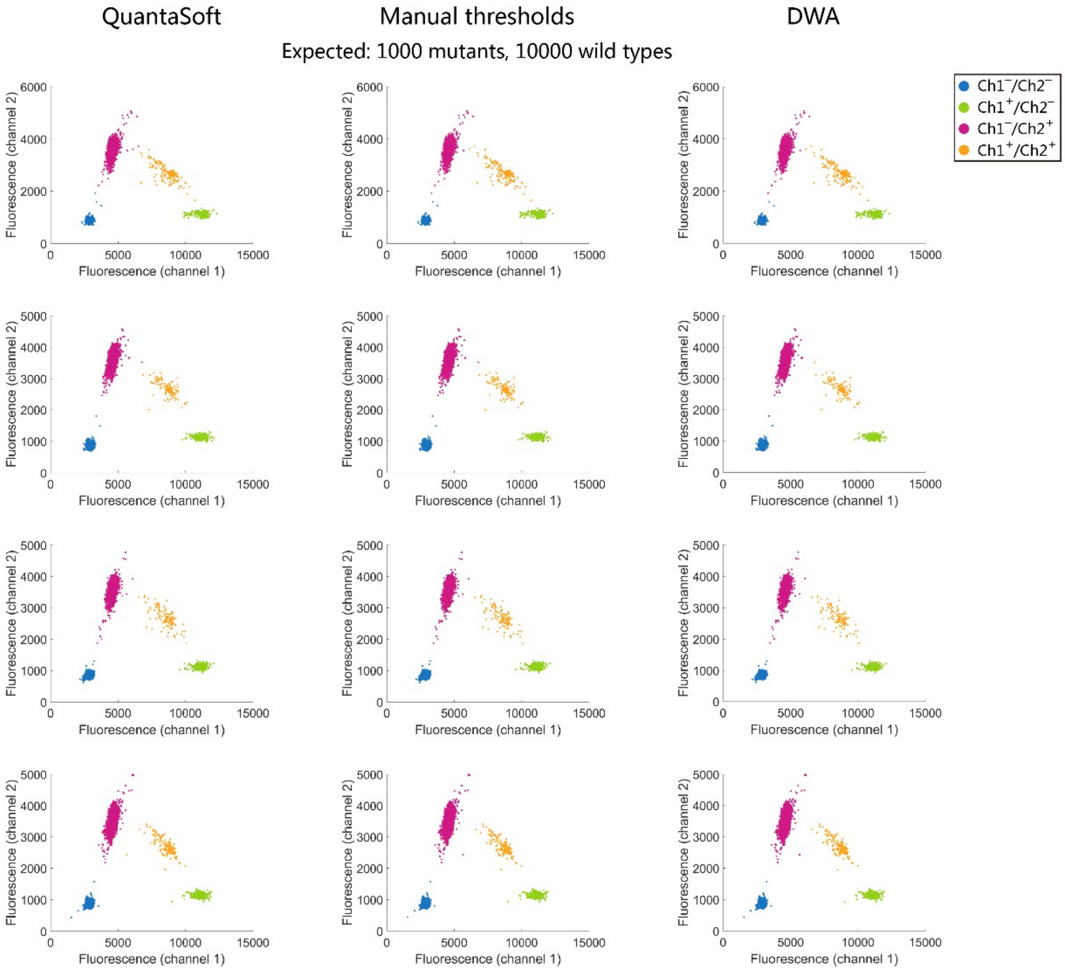


Figure S12. Comparison between classification results by QuantaSoft's automatic mode, manual thresholds and the DWA method. Each *EGFR* T790M assay contains about 1000 copies of mutants (Ch1) and about 10000 copies of wild types (Ch2). Each row is a replicate.

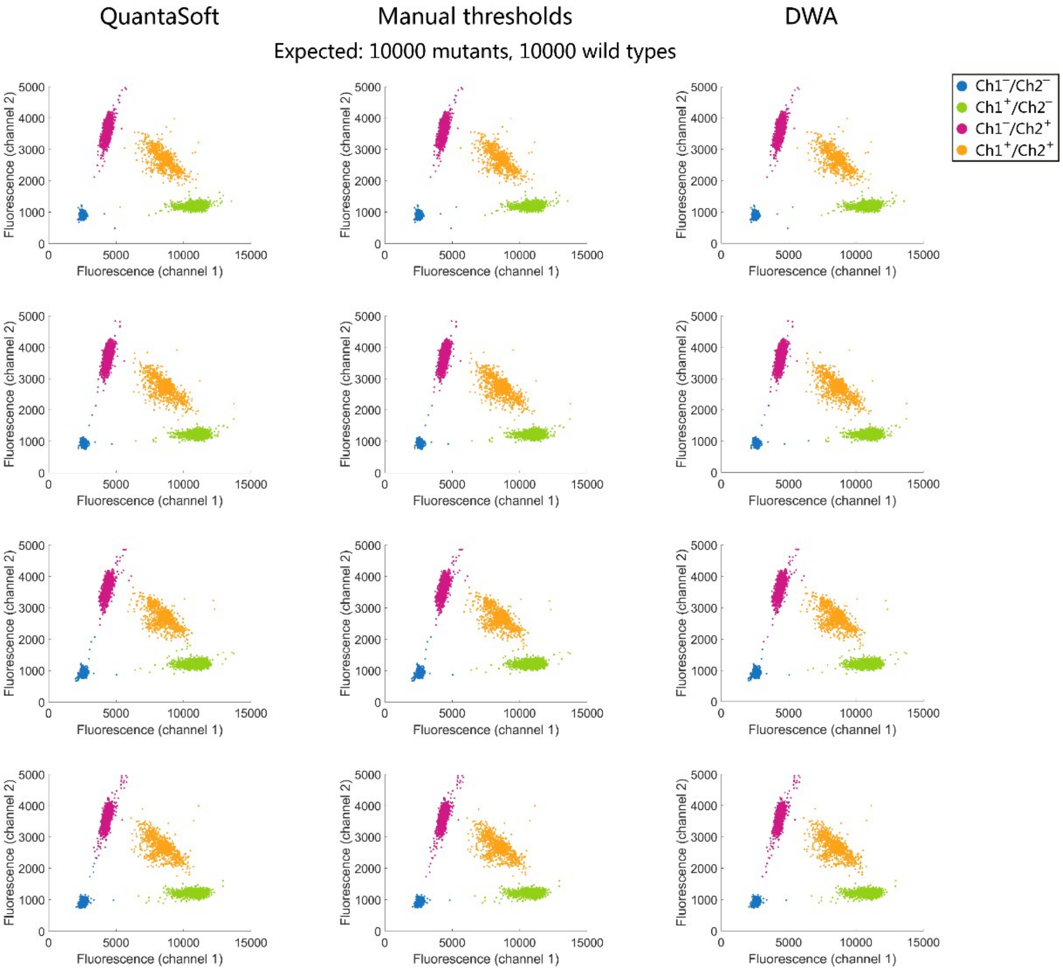


Figure S13. Comparison between classification results by QuantaSoft's automatic mode, manual thresholds and the DWA method. Each *EGFR* T790M assay contains about 10000 copies of mutants (Ch1) and about 10000 copies of wild types (Ch2). Each row is a replicate.

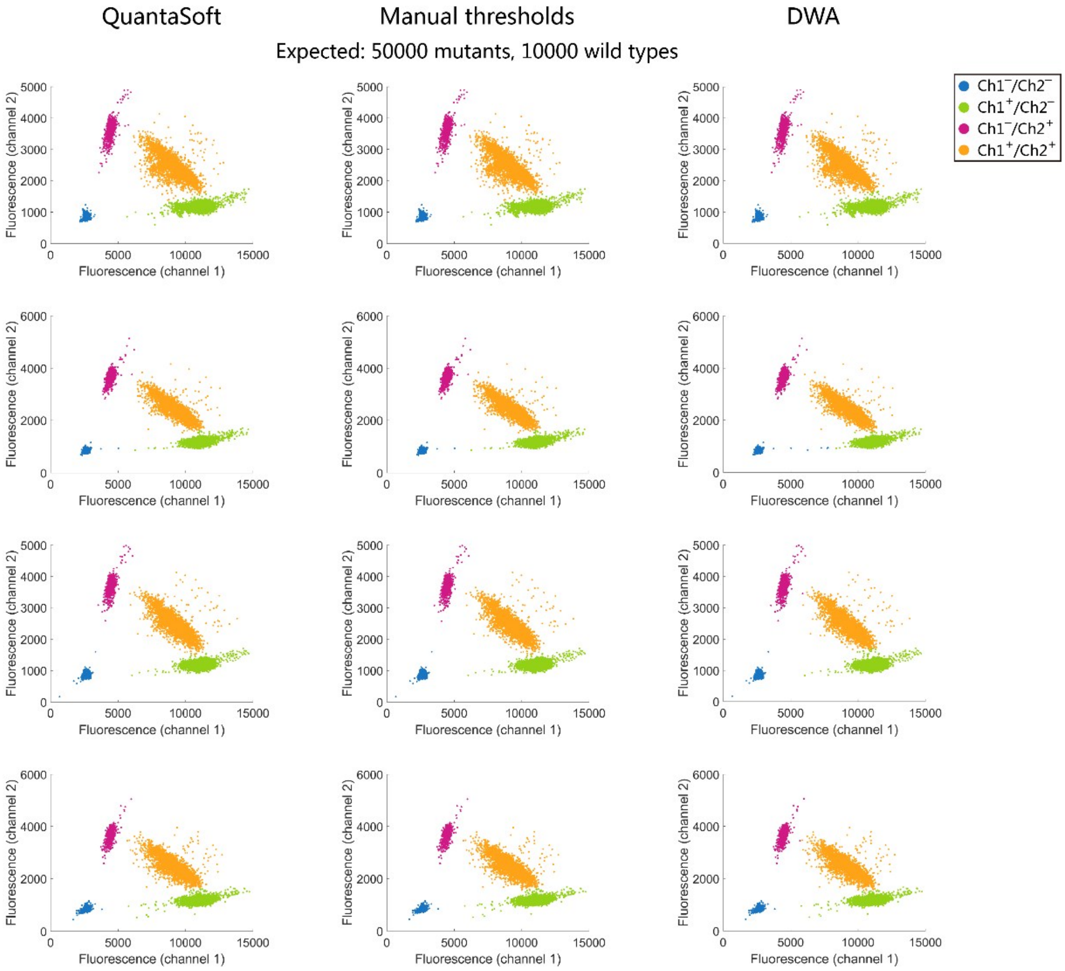


Figure S14. Comparison between classification results by QuantaSoft's automatic mode, manual thresholds and the DWA method. Each *EGFR* T790M assay contains about 50000 copies of mutants (Ch1) and about 10000 copies of wild types (Ch2). Each row is a replicate.

Supplementary Tables S1–4

Table S1. Statistical analysis results of the *EGFR* L858R classification results with the DWA method. (N: total droplet count, N_0: count of dual negative droplets, N_1: count of positive A/negative B droplets, N_2: count of negative A/positive B droplets, N_+: count of dual positive droplets, P-N ratio: positive-to-negative ratio)

Sample-replicate	N	Classification results				P-N ratio		Copy number	
		N_0	N_1	N_2	N_+	N_1/N_0	N_2/N_0	K_1	K_2
a-1	15605	10118	0	5487	0	0.00E+00	5.42E-01	0.0	10194.7
a-2	18631	11869	0	6762	0	0.00E+00	5.70E-01	0.0	10609.3
a-3	17866	11549	0	6317	0	0.00E+00	5.47E-01	0.0	10265.9
a-4	18216	11691	0	6525	0	0.00E+00	5.58E-01	0.0	10434.8
b-1	17111	11054	2	6055	0	1.81E-04	5.48E-01	4.3	10277.9
b-2	17296	11223	4	6069	0	3.56E-04	5.41E-01	8.4	10171.3
b-3	11814	7441	0	4373	0	0.00E+00	5.88E-01	0.0	10877.2
b-4	14386	9433	0	4953	0	0.00E+00	5.25E-01	0.0	9930.4
c-1	13092	8368	10	4710	4	1.20E-03	5.63E-01	28.1	10506.3
c-2	12924	8343	3	4577	1	3.60E-04	5.49E-01	8.5	10290.7
c-3	11176	7147	4	4021	4	5.60E-04	5.63E-01	13.2	10502.6
c-4	14338	9101	9	5225	3	9.89E-04	5.74E-01	23.3	10675.1
d-1	16824	10849	33	5920	22	3.04E-03	5.46E-01	71.5	10246.1
d-2	12942	8192	25	4710	15	3.05E-03	5.75E-01	71.7	10687.6
d-3	15486	9949	36	5488	13	3.62E-03	5.52E-01	85.0	10336.4
d-4	11036	7022	18	3992	4	2.56E-03	5.68E-01	60.2	10591.0
e-1	12709	8095	229	4252	133	2.83E-02	5.25E-01	656.4	9933.3
e-2	16985	10614	303	5895	173	2.85E-02	5.55E-01	662.3	10393.7
e-3	11193	6846	191	4052	104	2.79E-02	5.92E-01	647.5	10939.2
e-4	18135	11335	310	6312	178	2.73E-02	5.57E-01	634.9	10415.8
f-1	11980	6128	1670	3260	922	2.73E-01	5.32E-01	5670.6	10036.8
f-2	16781	8491	2319	4716	1255	2.73E-01	5.55E-01	5681.5	10393.9
f-3	11735	6008	1663	3130	934	2.77E-01	5.21E-01	5749.5	9867.1
f-4	16989	8648	2291	4699	1351	2.65E-01	5.43E-01	5529.6	10210.9
g-1	15301	2267	7566	1249	4219	3.34E+00	5.51E-01	34524.4	10326.3
g-2	14969	2198	7253	1302	4216	3.30E+00	5.92E-01	34319.4	10946.2
g-3	13795	2076	6782	1118	3819	3.27E+00	5.39E-01	34138.3	10137.2
g-4	16897	2472	8343	1447	4635	3.38E+00	5.85E-01	34727.2	10842.6

Table S2. Statistical analysis results of the *EGFR* T790M classification results with the DWA method. (N: total droplet count, N_0: count of dual negative droplets, N_1: count of positive A/negative B droplets, N_2: count of negative A/positive B droplets, N_+: count of dual positive droplets, P-N ratio: positive-to-negative ratio)

Sample-	N	Classification results				P-N ratio		Copy number	
replicate		N_0	N_1	N_2	N_+	N_1/N_0	N_2/N_0	K_1	K_2
a-1	14093	9610	0	4478	5	0.00E+00	4.66E-01	0.0	9000.5
a-2	14392	9816	0	4575	1	0.00E+00	4.66E-01	0.0	9002.1
a-3	15777	10683	0	5092	2	0.00E+00	4.77E-01	0.0	9171.1
a-4	15357	10360	0	4992	5	0.00E+00	4.82E-01	0.0	9254.0
b-1	15902	11074	3	4823	2	2.71E-04	4.36E-01	6.4	8506.6
b-2	15266	10311	1	4946	8	9.70E-05	4.80E-01	2.3	9219.5
b-3	17425	12080	2	5339	4	1.66E-04	4.42E-01	3.9	8612.0
b-4	15749	11229	2	4518	0	1.78E-04	4.02E-01	4.2	7956.5
c-1	13893	9558	7	4320	8	7.32E-04	4.52E-01	17.2	8774.7
c-2	15276	10685	15	4572	4	1.40E-03	4.28E-01	33.0	8381.1
c-3	16785	11454	9	5317	5	7.86E-04	4.64E-01	18.5	8972.1
c-4	18213	12230	10	5966	7	8.18E-04	4.88E-01	19.2	9348.5
d-1	14966	10153	34	4765	14	3.35E-03	4.69E-01	78.7	9054.1
d-2	15273	10329	28	4901	15	2.71E-03	4.74E-01	63.7	9136.7
d-3	15006	10163	28	4795	20	2.76E-03	4.72E-01	64.7	9093.9
d-4	17345	11682	32	5618	13	2.74E-03	4.81E-01	64.4	9239.0
e-1	16804	11015	273	5348	168	2.48E-02	4.86E-01	576.1	9312.1
e-2	16132	10635	260	5104	133	2.44E-02	4.80E-01	568.3	9223.3
e-3	14401	9599	229	4458	115	2.39E-02	4.64E-01	554.7	8975.6
e-4	16409	10700	260	5305	144	2.43E-02	4.96E-01	564.9	9474.3
f-1	11244	5818	1897	2607	922	3.26E-01	4.48E-01	6640.2	8711.7
f-2	12753	6362	2222	3073	1096	3.49E-01	4.83E-01	7048.4	9272.6
f-3	12962	6579	2268	3017	1098	3.45E-01	4.59E-01	6969.3	8881.5
f-4	13410	6766	2324	3159	1161	3.43E-01	4.67E-01	6947.4	9015.2
g-1	16671	2760	8522	1307	4082	3.09E+00	4.74E-01	33128.9	9121.8
g-2	14165	2349	7308	1087	3421	3.11E+00	4.63E-01	33263.4	8948.7
g-3	15169	2535	7772	1196	3666	3.07E+00	4.72E-01	33003.1	9093.7
g-4	13065	2182	6669	979	3235	3.06E+00	4.49E-01	32948.0	8721.1

Table S3. Comparison of QuantaSoft's automatic mode and the DWA method on two/four-ddPCR classification results of 117 clinical DNA samples derived from frozen tissues (FTs), formalin-fixed paraffin-embedded (FFPEs) tissues and peripheral blood (PB) for the detection of *EGFR* L858R and T790M wild types (WTs) and mutants (Muts). These ddPCR data were collected with Bio-Rad QX200 ddPCR system.

Sample index	Sample type	Target	Copy number: manual thresholds		Copy number error: QuantaSoft-manual		Copy number error: DWA-manual	
			WT	Mut	WT	Mut	WT	Mut
B101	FT	L858R	3773.6	12.7	-6.9	-9.1	0.0	0.0
B102	FT	L858R	3390.1	28.6	-10.5	-26.8	1.5	0.0
B103	FT	L858R	3608.7	18.1	-0.8	-16.3	0.0	0.0
B104	FT	L858R	3576.6	478.2	4.4	-471.0	2.1	1.8
B105	FT	L858R	4302.5	16.8	17.4	-16.8	0.0	0.0
B106	FT	L858R	3876.3	12.1	3.2	-10.1	2.0	0.0
B107	FT	L858R	5589.4	474.3	1.6	0.0	1.6	0.0
B108	FT	L858R	6270.9	29.1	35.0	-29.1	0.0	0.0
B109	FT	L858R	6486.5	31.2	37.9	-31.2	0.0	0.0
B110	FT	L858R	6388.4	1701.4	2.0	8.5	0.0	-2.0
B111	FT	L858R	4786.2	3386.4	-2.1	-0.3	-3.3	-3.0
B112	FT	L858R	22.2	14.3	-12.7	-14.3	0.0	6.4
B113	FT	L858R	9.4	9.4	9.4	0.0	0.0	0.0
B201	FT	T790M	1628.5	1.6	-1625.3	8.0	0.0	0.0
B202	FT	T790M	2445.9	0.0	-2436.8	45.2	0.0	0.0
B203	FT	T790M	4701.1	0.0	-4688.5	120.9	0.0	0.0
B204	FT	T790M	1492.6	121.1	-1488.4	-70.9	-4.2	0.0
B205	FT	T790M	1000.1	0.0	-996.5	67.7	-3.6	0.0
B206	FT	T790M	1648.3	0.0	-1641.9	229.7	0.0	0.0
B207	FT	T790M	1836.7	0.0	-1830.7	38.2	-1.0	0.0
B208	FT	T790M	513.9	1.4	-513.9	80.1	0.0	0.0
B209	FT	T790M	1429.3	0.0	-1418.8	44.1	0.0	0.0
B210	FT	T790M	817.2	1.6	-809.4	96.8	0.0	0.0
B211	FT	T790M	2324.1	4.5	-2315.0	0.0	-2.0	0.0
B212	FT	T790M	641.2	2.2	-373.5	-2.2	0.0	0.0
B213	FT	T790M	1262.1	0.0	-1259.3	75.8	0.0	0.0
B214	FT	T790M	1001.3	0.0	-990.3	83.6	3.5	0.0
B215	FT	T790M	2247.5	0.0	-2247.5	68.0	-2.6	0.0
B216	FT	T790M	1980.0	137.8	0.0	0.0	0.0	0.0
B217	FT	T790M	1687.7	160.3	-1.2	0.0	-0.3	0.0

Sample index	Sample type	Target	Copy number: manual thresholds		Copy number error: QuantaSoft-manual		Copy number error: DWA-manual	
			WT	Mut	WT	Mut	WT	Mut
B301	FFPE	L858R	1917.7	774.5	-2.0	70.8	12.6	92.6
B302	FFPE	L858R	1336.1	972.9	58.5	66.8	90.6	166.9
B303	FFPE	L858R	2057.2	0.0	-1882.6	31.6	4.2	0.0
B304	FFPE	L858R	1601.9	2573.5	88.6	30.7	75.5	-158.5
B305	FFPE	L858R	2954.1	1119.8	-48.9	2.6	169.8	11.7
B306	FFPE	L858R	10772.8	14.8	-10430.5	11.9	-15.4	4.9
B307	FFPE	L858R	581.7	1470.9	44.8	108.1	-6.8	-38.7
B308	FFPE	L858R	9413.1	8.6	-9109.7	21.3	-80.6	2.8
B309	FFPE	L858R	6840.9	2565.0	-126.1	-21.3	-27.7	-18.7
B310	FFPE	L858R	1945.7	2.1	-705.3	0.0	48.1	0.0
B311	FFPE	L858R	1193.7	1934.9	40.2	12.5	0.9	-63.7
B401	FFPE	T790M	11251.2	17.8	-11120.9	23.8	83.1	0.1
B402	FFPE	T790M	6890.8	13.5	-6789.6	60.4	-10.8	0.0
B403	FFPE	T790M	15244.7	25.6	-14895.2	63.1	-90.2	-3.3
B404	FFPE	T790M	5743.8	18.5	-5633.6	49.6	-45.8	-9.3
B405	FFPE	T790M	11682.3	14.8	-11589.9	59.7	0.0	0.0
B406	FFPE	T790M	5082.3	11.9	-4966.7	38.8	128.5	0.1
B407	FFPE	T790M	17886.2	24.4	-17778.8	218.7	31.3	4.1
B408	FFPE	T790M	26989.8	27.1	-26808.3	225.6	155.3	11.1
B409	FFPE	T790M	16769.9	17.3	-16578.8	101.1	15.4	0.0
B410	FFPE	T790M	7756.7	16.6	-7643.1	20.2	-110.2	9.8
B411	FFPE	T790M	5824.7	18.0	-5705.6	71.4	155.1	0.1
B412	FFPE	T790M	4503.1	11.4	-4380.4	48.8	-138.3	2.2
B413	FFPE	T790M	3291.1	13.2	-178.9	0.0	77.2	0.0
B414	FFPE	T790M	6276.4	15.5	-6098.7	43.8	-136.4	6.1
B415	FFPE	T790M	5238.9	9.6	-5178.5	26.6	-24.1	2.4
B416	FFPE	T790M	7011.5	11.6	-6927.4	84.2	154.9	-2.3
B417	FFPE	T790M	12267.3	31.2	-12084.4	69.9	90.7	0.1
B501	PB	L858R	12236.3	0.0	194.1	0.0	0.0	0.0
B502	PB	L858R	8945.8	0.0	250.4	0.0	74.8	0.0
B503	PB	L858R	9758.1	125.6	154.2	-94.2	-46.7	3.9
B504	PB	L858R	11251.7	85.0	-81.4	-0.3	60.9	0.2
B505	PB	L858R	8514.2	177.9	-116.4	-0.9	5.0	-2.4
B506	PB	L858R	9571.0	0.0	193.1	0.0	41.7	0.0
B507	PB	L858R	9529.4	0.0	180.0	0.0	-55.8	0.0
B508	PB	L858R	10841.7	0.0	151.9	0.0	10.8	0.0
B509	PB	L858R	8857.5	71.1	-4.1	0.0	44.1	-1.9
B510	PB	L858R	10439.6	4.1	243.8	-4.1	0.0	0.0
B511	PB	L858R	1044.0	0.0	-1039.1	6.1	3.7	0.0
B512	PB	L858R	10361.7	4.0	114.2	-4.0	-81.9	0.0
B513	PB	L858R	10331.1	0.0	4.1	0.0	18.5	0.0

Sample index	Sample type	Target	Copy number: manual thresholds		Copy number error: QuantaSoft-manual		Copy number error: DWA-manual	
			WT	Mut	WT	Mut	WT	Mut
B514	PB	L858R	13175.0	25.3	17.7	-25.3	2.5	0.0
B515	PB	L858R	12020.9	0.0	114.8	0.0	-47.7	0.0
B516	PB	L858R	10980.2	0.0	90.3	0.0	-83.3	0.0
B517	PB	L858R	10120.9	162.4	378.7	-162.4	77.9	2.7
B518	PB	L858R	9252.1	176.8	-9252.1	-176.8	81.9	0.6
B519	PB	L858R	10357.2	0.0	121.8	0.0	36.5	0.0
B520	PB	L858R	9975.7	0.0	258.4	0.0	-47.3	0.0
B521	PB	L858R	7563.2	187.9	325.9	-187.9	5.3	0.0
B522	PB	L858R	11569.7	0.0	-2.9	0.0	5.9	0.0
B601	PB	T790M	8348.4	0.0	-8259.0	16.4	10.8	0.0
B602	PB	T790M	7414.7	6.3	-7362.9	3.7	-2.1	-2.1
B603	PB	T790M	9191.1	0.0	-9153.0	8.5	-44.3	0.0
B604	PB	T790M	7861.4	0.0	-7821.2	4.7	-110.5	0.0
B605	PB	T790M	10655.4	2.3	-10610.5	14.2	102.8	2.4
B606	PB	T790M	7395.2	6.1	-7321.5	2.1	-16.3	2.0
B607	PB	T790M	6923.1	0.0	-6892.1	13.8	51.9	0.0
B608	PB	T790M	7885.7	2.2	-7855.3	13.1	26.0	0.0
B609	PB	T790M	9832.8	2.0	-9794.4	8.1	12.7	4.0
B610	PB	T790M	7194.5	0.0	-7159.0	18.8	4.2	0.0
B611	PB	T790M	9525.8	0.0	162.8	0.0	43.7	0.0
B612	PB	T790M	7983.5	3.4	-7961.2	-1.7	-42.8	0.0
B613	PB	T790M	7639.6	1.8	-7621.8	12.5	-113.7	0.0
B614	PB	T790M	7728.1	0.0	-7680.2	5.1	-59.9	0.0
B615	PB	T790M	7002.0	0.0	-6990.4	3.9	-54.0	0.0
B616	PB	T790M	8349.9	6.0	-8303.7	12.1	36.4	0.0
B617	PB	T790M	8452.2	0.0	-8391.5	10.9	-160.3	0.0
B618	PB	T790M	8187.9	28.2	-8145.5	8.1	-76.6	-0.1
B619	PB	T790M	8090.2	1.8	-8044.1	21.3	-47.9	0.0
B620	PB	T790M	6427.2	0.0	-6402.6	34.5	-81.7	0.0
B621	PB	T790M	6711.1	0.0	-6684.6	24.9	10.1	0.0
B622	PB	T790M	4909.2	0.0	-4909.2	0.0	35.5	0.0
B623	PB	T790M	6827.8	3.6	-6725.5	50.3	75.1	0.0
B624	PB	T790M	6629.3	1.6	-6602.7	1.6	-13.3	0.0
B625	PB	T790M	7652.0	3.6	-7625.1	0.0	-41.1	0.0
B626	PB	T790M	9184.9	1.8	-9148.8	14.4	14.4	0.0
B627	PB	T790M	10206.5	4.0	-10176.0	6.1	-72.6	0.0
B628	PB	T790M	8325.9	1.8	-8289.8	12.6	-43.0	0.0
B629	PB	T790M	8787.4	1.9	-8741.7	1.9	97.4	0.0
B630	PB	T790M	10337.7	4.4	-10324.4	2.2	-55.5	0.0
B631	PB	T790M	9110.4	2.2	-9103.8	8.8	-87.7	0.0
B632	PB	T790M	6919.7	127.7	86.8	0.5	-24.7	1.9

Sample index	Sample type	Target	Copy number: manual thresholds		Copy number error: QuantaSoft-manual		Copy number error: DWA-manual	
			WT	Mut	WT	Mut	WT	Mut
B633	PB	T790M	5748.4	0.0	-5724.1	17.0	-50.8	0.0
B634	PB	T790M	7906.0	5.7	-7877.2	5.8	-71.4	0.0
B635	PB	T790M	8611.3	0.0	-8611.3	0.0	38.1	0.0
B636	PB	T790M	9838.7	2.7	-9819.5	5.5	0.0	0.0
B637	PB	T790M	13488.4	23.0	-13462.7	2.7	77.0	0.1

Table S4. Validation the DWA method on two/four-ddPCR classification results of 137 clinical DNA samples derived from frozen tissues (FTs), formalin-fixed paraffin-embedded (FFPEs) tissues and peripheral blood (PB) for the detection of *EGFR* L858R and T790M wild types (WTs) and mutants (Muts). These ddPCR data were collected with TargetingOne TD-1 ddPCR system.

Sample index	Sample type	Target	Copy number: manual thresholds		Copy number error: DWA-manual	
			WT	Mut	WT	Mut
T101	FT	L858R	3327.5	4.0	0.0	0.0
T102	FT	L858R	4173.9	0.0	0.0	0.0
T103	FT	L858R	7678.6	971.8	0.0	0.0
T104	FT	L858R	55372.4	63832.3	0.0	0.0
T105	FT	L858R	4077.6	294.4	0.0	8.6
T106	FT	L858R	2981.3	0.0	0.0	0.0
T107	FT	L858R	3413.6	0.0	0.0	0.0
T108	FT	L858R	7858.7	0.0	0.0	0.0
T109	FT	L858R	4472.2	144.4	-0.1	-2.2
T110	FT	L858R	1653.2	440.5	0.0	0.0
T111	FT	L858R	3624.1	3.4	0.1	1.7
T112	FT	L858R	1614.4	47.4	0.0	0.0
T113	FT	L858R	3144.0	106.0	0.0	0.0
T114	FT	L858R	2407.2	3884.9	0.0	1.5
T115	FT	L858R	874.2	0.0	0.0	0.0
T201	FT	T790M	575.1	0.0	0.0	0.0
T202	FT	T790M	823.9	0.0	1.6	0.0
T203	FT	T790M	2560.4	0.0	0.0	0.0
T204	FT	T790M	1122.8	1.3	0.0	0.0
T205	FT	T790M	4041.7	0.0	0.0	0.0
T206	FT	T790M	2214.1	1.8	-0.1	0.0
T207	FT	T790M	4082.7	3.8	-0.1	0.0
T208	FT	T790M	1847.0	0.0	0.0	0.0
T209	FT	T790M	1254.1	0.0	0.0	0.0
T210	FT	T790M	2014.8	0.0	0.0	0.0
T211	FT	T790M	873.6	2.1	0.0	0.0
T212	FT	T790M	1440.9	3.8	-1.8	0.0
T213	FT	T790M	1248.4	1.3	0.0	0.0
T301	FFPE	L858R	926.1	57.8	0.0	0.0
T302	FFPE	L858R	1674.5	229.4	-1.7	-1.7

Sample index	Sample type	Target	Copy number: manual thresholds		Copy number error: DWA-manual	
			WT	Mut	WT	Mut
T303	FFPE	L858R	1739.9	231.0	0.0	0.0
T304	FFPE	L858R	1617.2	37.6	3.3	0.0
T305	FFPE	L858R	11294.9	31.9	0.0	0.0
T306	FFPE	L858R	4471.9	1216.3	0.0	0.0
T307	FFPE	L858R	183.6	42.9	0.0	-2.7
T308	FFPE	L858R	2017.1	45.4	0.0	0.0
T309	FFPE	L858R	5435.2	2095.5	0.0	0.0
T310	FFPE	L858R	4663.1	42.6	0.0	0.0
T311	FFPE	L858R	682.6	30.9	0.0	0.0
T312	FFPE	L858R	1446.7	35.5	-0.1	-3.4
T313	FFPE	L858R	4043.9	773.2	1.4	-1.3
T314	FFPE	L858R	3454.0	27.7	0.0	0.0
T315	FFPE	L858R	1596.1	353.1	1.6	0.0
T316	FFPE	L858R	4760.2	319.9	0.0	0.0
T317	FFPE	L858R	1704.4	35.5	0.0	0.0
T318	FFPE	L858R	2044.7	40.4	3.2	0.0
T319	FFPE	L858R	749.4	48.7	0.0	0.0
T320	FFPE	L858R	1229.4	22.8	-0.6	-1.8
T321	FFPE	L858R	3880.3	606.0	0.0	-1.9
T322	FFPE	L858R	2212.5	76.5	1.4	0.0
T323	FFPE	L858R	4194.4	29.9	0.0	0.0
T324	FFPE	L858R	2947.5	103.0	-3.3	0.0
T325	FFPE	L858R	4152.6	9.0	-4.5	0.0
T326	FFPE	L858R	5583.9	2.1	0.0	0.0
T327	FFPE	L858R	4657.6	364.3	3.9	0.0
T328	FFPE	L858R	6812.8	2.1	0.0	0.0
T329	FFPE	L858R	7817.4	2.1	0.0	0.0
T330	FFPE	L858R	1662.9	4.1	0.0	0.0
T331	FFPE	L858R	3476.2	0.0	0.0	0.0
T332	FFPE	L858R	1638.6	2.0	0.0	0.0
T333	FFPE	L858R	3639.5	1516.3	0.0	0.0
T334	FFPE	L858R	2946.7	3.1	0.0	0.0
T335	FFPE	L858R	4374.6	372.6	-1.9	0.0
T336	FFPE	L858R	14238.6	1300.0	0.0	0.0
T337	FFPE	L858R	1276.7	2.0	-3.8	0.0
T338	FFPE	L858R	706.8	9.4	0.0	0.0
T339	FFPE	L858R	5593.9	1921.9	0.0	-2.0
T401	FFPE	T790M	4800.1	0.0	0.0	0.0
T402	FFPE	T790M	4587.7	0.0	0.0	0.0
T403	FFPE	T790M	1269.1	4.3	0.0	0.0
T404	FFPE	T790M	3756.9	0.0	0.0	0.0

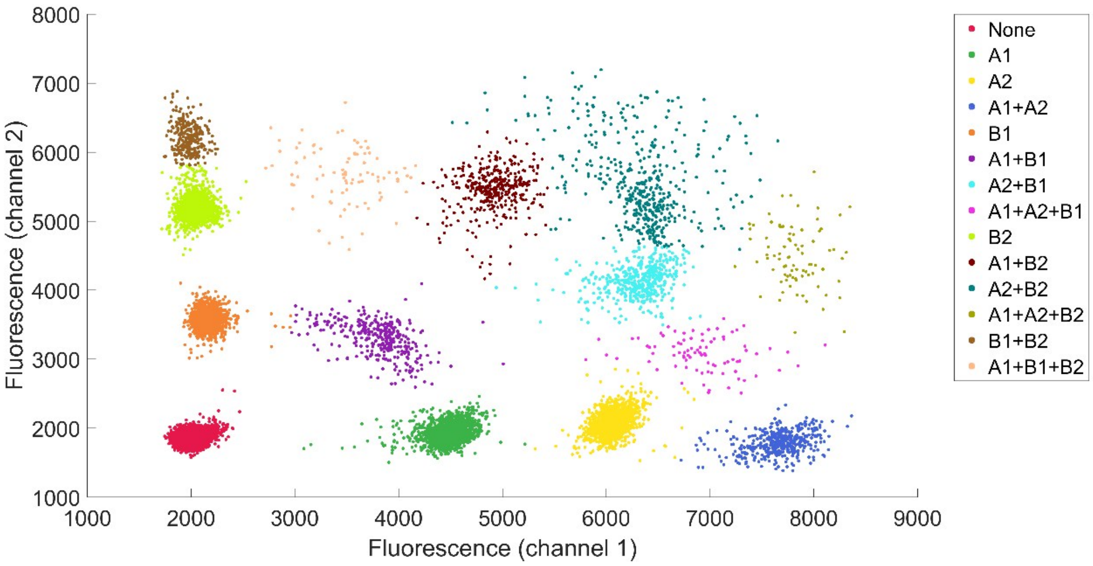
Sample index	Sample type	Target	Copy number: manual thresholds		Copy number error: DWA-manual	
			WT	Mut	WT	Mut
T405	FFPE	T790M	14640.9	0.0	0.0	0.0
T406	FFPE	T790M	3771.0	1.2	0.0	0.0
T407	FFPE	T790M	4696.1	0.0	0.0	0.0
T408	FFPE	T790M	767.8	0.0	0.0	0.0
T409	FFPE	T790M	7536.4	0.0	-6.1	0.0
T410	FFPE	T790M	4655.6	5.1	-2.1	-3.4
T411	FFPE	T790M	1272.0	1.5	0.0	0.0
T412	FFPE	T790M	1201.0	1.2	0.0	0.0
T413	FFPE	T790M	1059.9	0.0	0.0	0.0
T414	FFPE	T790M	816.7	2.1	0.0	0.0
T415	FFPE	T790M	952.1	0.0	0.0	0.0
T416	FFPE	T790M	1548.1	0.0	-1.3	0.0
T417	FFPE	T790M	3769.1	1.6	-6.3	0.0
T418	FFPE	T790M	4634.4	0.0	0.0	0.0
T419	FFPE	T790M	4425.0	0.0	2.0	0.0
T420	FFPE	T790M	343.7	0.0	0.0	0.0
T421	FFPE	T790M	5172.4	1.9	-1.9	0.0
T422	FFPE	T790M	306.9	0.0	0.0	0.0
T423	FFPE	T790M	2284.3	0.0	1.2	0.0
T424	FFPE	T790M	5192.1	0.0	-3.7	0.0
T425	FFPE	T790M	4004.0	0.0	0.0	0.0
T426	FFPE	T790M	4109.6	0.0	0.0	1.4
T427	FFPE	T790M	4120.2	1.5	0.0	0.0
T428	FFPE	T790M	1132.2	0.0	0.0	0.0
T429	FFPE	T790M	4290.4	0.0	0.0	0.0
T430	FFPE	T790M	4484.8	0.0	0.0	0.0
T431	FFPE	T790M	1116.5	0.0	0.0	0.0
T432	FFPE	T790M	3794.1	0.0	-1.5	0.0
T433	FFPE	T790M	2797.1	1.9	0.0	0.0
T434	FFPE	T790M	3987.9	2.0	0.0	0.0
T435	FFPE	T790M	1853.9	0.0	-4.9	0.0
T436	FFPE	T790M	4210.9	0.0	2.7	0.0
T437	FFPE	T790M	3406.0	0.0	0.0	0.0
T438	FFPE	T790M	5103.5	0.0	0.0	0.0
T439	FFPE	T790M	1209.6	0.0	0.0	0.0
T440	FFPE	T790M	5704.4	0.0	1.8	0.0
T441	FFPE	T790M	7548.8	0.0	1.7	0.0
T442	FFPE	T790M	2353.9	0.0	0.0	0.0
T443	FFPE	T790M	1801.5	0.0	3.9	0.0
T444	FFPE	T790M	5153.1	0.0	-2.1	0.0
T501	PB	L858R	12164.6	0.0	0.0	0.0

Sample index	Sample type	Target	Copy number: manual thresholds		Copy number error: DWA-manual	
			WT	Mut	WT	Mut
T502	PB	L858R	11788.0	26.8	0.4	2.4
T503	PB	L858R	10031.0	85.1	0.0	0.0
T504	PB	L858R	9382.4	0.0	-4.0	0.0
T505	PB	L858R	8930.9	0.0	0.0	0.0
T506	PB	L858R	9079.3	0.0	-1.2	0.0
T507	PB	L858R	9737.0	0.0	-5.9	0.0
T508	PB	L858R	7507.0	0.0	1.5	0.0
T509	PB	L858R	9722.8	3.0	0.0	0.0
T510	PB	L858R	12927.1	0.0	6.2	0.0
T511	PB	L858R	9780.3	0.0	-1.7	0.0
T601	PB	T790M	6766.5	143.1	2.0	-4.1
T602	PB	T790M	9694.9	0.0	0.0	0.0
T603	PB	T790M	8847.4	0.0	0.0	0.0
T604	PB	T790M	9076.9	2.2	0.3	0.0
T605	PB	T790M	10053.7	0.0	0.0	0.0
T606	PB	T790M	26320.2	7.5	-1.3	0.0
T607	PB	T790M	7974.2	3.8	-3.8	0.0
T608	PB	T790M	15755.8	1.9	0.0	0.0
T609	PB	T790M	8508.9	2.6	-1.1	1.3
T610	PB	T790M	10237.7	2.2	0.0	0.0
T611	PB	T790M	8672.5	5.5	0.0	0.0
T612	PB	T790M	7187.5	0.0	0.0	0.0
T613	PB	T790M	7784.8	2.9	0.0	0.0
T614	PB	T790M	12677.8	0.0	0.0	0.0
T615	PB	T790M	8973.3	1.7	0.0	0.0

Supplementary Methods: Mathematical Description of Density-Watershed Algorithm (DWA) Method

To facilitate automatic classification of ddPCR data with customized software, we provide a mathematical description of the density-watershed algorithm in this section. To be noted, except for the illustration of particular cases, this mathematic description of the DWA method applies to ddPCR data of any known dimension that can be classified into at most any known number of clusters, which is not limited to two/four-ddPCR data.

For example, the DWA method can be used to classify the following dual-fluorescence and sixteen-cluster ddPCR data (two/sixteen-ddPCR data), typically derived from multiplexing two targets for each of the two fluorescence channel (A1, A2 and B1, B2) through an amplitude-based way [1].



In this two/sixteen-ddPCR data set, some clusters are well-defined, such as none, A1, A2, B1, B2, A1+B1, and A2+B2. These well-defined clusters are correctly classified and identified with the DWA method. The rest (which usually contain three or more kinds of targets) are poorly-defined due to misalignment or murky rain data. Even with misalignment and rain data, these clusters are reasonably classified and identified with the DWA method. This demonstrates the capability of the DWA method in classifying multiplexed ddPCR data.

A. Step 1: data gridding of ddPCR fluorescence scatter plot

Appropriate data gridding is performed in a self-adaptive way to calculate data densities in a ddPCR data scatter plot. The key to reasonable data gridding is the determination of grid numbers. As previously described, the data distribution in each cluster conforms (multivariate) normal distribution [2]. Normal distribution can be approximated with symmetric binomial distribution (with success probability as 0.5), and the number of grids in a binomial distribution is logarithmically correlated with data count (number of trials). In this way, Sturges' formula can be applied to calculate the number of grids on each dimension of ddPCR data [3]. We assume ddPCR data to be $d_i = \left(d_{i1}, d_{i2}, \cdots, d_{idim(d_i)}\right)^T, i = 1,2,\cdots,n$. A set D_j for the data components on each dimension is defined as $D_j \triangleq \{d_{ij}\}$, and then the number of grids on that dimension can be calculated with equation (1).

$$s_j \triangleq \lceil log_2[card(D_j)] \rceil + 1 \tag{1}$$

B. Step 2: density-based watershed algorithm

The density-based watershed algorithm uses data counts in each grid as data density and automatically segments the gridded scatter plot into isolated regions based on the data densities. Data densities ρ in each grid is defined as the data count in that grid. After that, the DWA method needs to find a series of border grids G^* to segment the gridded scatter plot into regions R_i, while minimizing the total data counts in these border grids G^* subject to the following criteria as mathematically displayed in equation (2): (a) Each region R_i contains one or more inner grids G_j; (b) The inner grids G_j in different regions R_i do not share any geometric features (such as edges and vertices) in common; (c) Each local maxima of data density correspond to inner grids of different regions.

$$\begin{gathered} \bigcup G^* = \underset{G^*}{\operatorname{argmin}} \sum_{G^*} \rho \\ s.t. \begin{cases} \forall G_{j_1} \subseteq R_{i_1}, G_{j_2} \subseteq R_{i_2}, j_1 \neq j_2, i_1 \neq i_2; G_{j_1} \cap G_{j_2} = \emptyset \\ \forall \underset{G_{j_1} \subseteq R_{i_1}}{\operatorname{argmax}} \rho \neq \underset{G_{j_2} \subseteq R_{i_2}}{\operatorname{argmax}} \rho, j_1 \neq j_2; i_1 \neq i_2 \end{cases} \end{gathered} \tag{2}$$

Practically, density-based watershed algorithm can be applied to automatically find these border grids conforming equation (2) [4, 5]. First, a density terrain is constructed with $-\rho$, the opposites of data densities. Then, flooding-based watershed algorithms [6] is applied with $3^{dim(d_i)} - 1$ connectivity. Particularly, for 1-D ddPCR data, each grid is connected to two grids except for those at the ends; for 2-D ddPCR data, each grid is connected to eight grids except for those on the edges or at the vertices.

C. Step 3: determination of optimal cluster pattern

For these isolated regions, optimal cluster pattern can be automatically determined for each two/four-ddPCR data according to these regions' distribution without supervised learning.

In the first step, for each possible cluster pattern $k_1 \times k_2 \times \cdots \times k_i \times \cdots k_{dim(d_i)}$, benchmarks can be positioned at every possible combination of the coordinates in equation (3) as $P_j\left(x_1, x_2, \cdots, x_i, \cdots, x_{dim(d_i)}\right)$, where k_i is the number of unique clusters when all clusters in the current cluster pattern is projected to dimension i.

$$x_i = \begin{cases} \frac{1}{2}min(D_i) + \frac{1}{2}max(D_i), k_i = 1 \\ \frac{k_i - a}{k_i - 1}min(D_i) + \frac{a - 1}{k_i - 1}max(D_i), a = 1, 2, \cdots, k_i, k_i \geq 2 \end{cases} \quad (3)$$

Particularly, for 2/4-ddPCR data (target template A for Ch1 and target template B for Ch2), four possible cluster patterns are listed as: (a) Ch1$^-$/ Ch2$^-$, one cluster (dual negative); (b) Ch1$^+$/ Ch2$^-$, two clusters (dual negative and positive A/negative B); (c) Ch1$^-$/ Ch2$^+$, two clusters (dual negative and negative A/positive B); and (d) Ch1$^+$/ Ch2$^+$, three or four clusters (dual negative, positive A/negative B, negative A/positive B, and optional dual positive). In this way, benchmark(s) can be positioned for each possible cluster pattern as follows:

(a) Ch1$^-$/ Ch2$^-$: a single benchmark is positioned at P_1 as shown in equation (4);

$$P_1\left(\frac{min(D_1) + max(D_1)}{2}, \frac{min(D_2) + max(D_2)}{2}\right) \quad (4)$$

(b) Ch1$^+$/ Ch2$^-$: two benchmarks are positioned at P_1 and P_2 as shown in equation (5);

$$P_1\left(min(D_1), \frac{min(D_2) + max(D_2)}{2}\right) \qquad P_2\left(max(D_1), \frac{min(D_2) + max(D_2)}{2}\right) \quad (5)$$

(c) Ch1$^-$/ Ch2$^+$: two benchmarks are positioned at P_1 and P_2 as shown in equation (6);

$$\begin{gathered} P_2\left(\frac{min(D_1) + max(D_1)}{2}, max(D_2)\right) \\ P_1\left(\frac{min(D_1) + max(D_1)}{2}, min(D_2)\right) \end{gathered} \quad (6)$$

(d) Ch1⁺/ Ch2⁺: four benchmarks are positioned at P_1, P_2, P_3 and P_4 as shown in equation (7).

$$\begin{matrix} P_3\left(min\left(D_1\right),max\left(D_2\right)\right) & P_4\left(max\left(D_1\right),max\left(D_2\right)\right) \\ P_1\left(min\left(D_1\right),min\left(D_2\right)\right) & P_2\left(max\left(D_1\right),min\left(D_2\right)\right) \end{matrix} \tag{7}$$

In the second step, for each possible cluster pattern, DWA calculates the distance of each region center (the centroid weighed by data counts in each inner grid) to the nearest benchmark position. Weighed by total data counts q_i in each region R_i, DWA sums up all the distances into a total nearest distance (TND) T, as shown in equation (8).

$$T \triangleq \sum_i \left(q_i \min_j \|C_i - P_j\| \right) \tag{8}$$

In the last step, DWA designates the optimal cluster pattern as the one that yields minimal TND, as shown in equation (9). If there is more than one set of $\left(k_1, k_2,\cdots,k_i,\cdots,k_{dim\left(d_i\right)}\right)$ that yields the minimal TND, the set with the minimal sum $\sum_i k_i$ is used.

$$\left(k_1, k_2,\cdots,k_i,\cdots,k_{dim\left(d_i\right)}\right) = \underset{k}{\operatorname{argmin}}\, T \tag{9}$$

D. Step 4: selection and merging of regions

After the determination of optimal cluster pattern, the DWA method is required to select regions to represent the clusters in the optimal pattern and then merge the unselected regions automatically after two rounds of region selection and one round of region merging.

During the first round of region selection (two-way selection), only region centers and benchmark positions that are the nearest neighbor to each other are paired as $\left(C_{i^*}, P_{j^*}\right)$ and labeled as selected, as shown in equation (10). During the second round of region selection (one-way selection), an unselected region center $C_{i^\times}$ and an unselected benchmark position $P_{j^\times}$ were selected and paired when the unselected region center $C_{i^\times}$ was the nearest neighbor to the unselected benchmark position $P_{j^\times}$, yielding $\left(C_{i^\times}, P_{j^\times}\right)$ as shown in equation (11). The corresponding benchmark position and region center are marked as selected after the pairing.

$$\left(C_{i^*}, P_{j^*}\right) \in \left\{\left(C_{i^*}, P_{j^*}\right) \middle| i^* = \underset{i}{\operatorname{argmin}} \left\|C_i - P_{j^*}\right\|, j^* = \underset{j}{\operatorname{argmin}} \left\|C_{i^*} - P_j\right\|\right\} \tag{10}$$

$$\left(C_{i^\times}, P_{j^\times}\right) \in \{\left(C_{i^\times}, P_{j^\times}\right) \big| i^\times = \underset{i}{\operatorname{argmin}} \left\|C_i - P_{j^\times}\right\|, C_{i^\times} \notin \left\{C_{i^*}\right\}, P_{j^\times} \notin \{P_{j^*} \tag{11}$$

During region merging, DWA merges each unselected region (whose center does not belong to $\left\{C_{i^*}\right\} \cup \left\{C_{i^\times}\right\}$) into another region with the nearest center. The border grids adjacent only to the merging regions are converted to inner grids of the merged region, and the center of the merged region is recalculated as the centroid weighed by data counts in the updated inner grids. If one of the merging regions is labeled as selected, the center of the merged region is labeled as selected, otherwise it is not. This process is repeated until all regions are selected.

E. Step 5: classification of ddPCR data

Based on the merged regions, the DWA method was able to create clusters for each region and classify ddPCR data into the clusters. Each merged region represents a cluster in classification result. The classification process consists three steps.

In the first step, the merged regions are aligned with the ddPCR data scatter plot.

The outer boundaries in each dimension of the regions are aligned with the boundaries in each dimension of the ddPCR data in the scatter plot.

In the second step, all ddPCR data that fall in inner grids of the same region are classified into the same cluster, since each merged region represents a different cluster.

In the third step, the rest of ddPCR data, which fall in border grids, are classified according to Bayesian criterion on adjacent regions. This Bayesian criterion is applied as follows. First, (multivariate) normal distribution $N(\hat{\mu}_i,\hat{\Sigma}_i)$ is constructed for each formed cluster with estimated mean $\hat{\mu}_i$ and (co)variance $\hat{\Sigma}_i$. By assuming hypodispersion of data within each grid, the parameters $\hat{\mu}_i$ and $\hat{\Sigma}_i$ can be estimated with region parameters and grid parameters instead of direct computation from large amounts of ddPCR data, as shown in equation (12) and (13), in which region parameters for R_i include center coordinate c_i and data count q_i, and grid parameters for G_j include center coordinate φ_j, data density ρ_j and edge lengths $\Delta x_1,\Delta x_2,\cdots,\Delta x_{dim(d_i)}$. Then, Bayesian discriminant [7] is applied to find the region index i_k^* that yields maximal posteriori probability of each data point among the regions adjacent to the border grid where the data point in question is located, as shown in equation (14).

$$\begin{aligned}\hat{\mu}_i &= \frac{1}{q_i}\sum_{G_j \subseteq R_i}\left\{\frac{\rho_j}{\prod_k \Delta x_k}\int_{-\frac{\Delta x_{dim(d_i)}}{2}}^{\frac{\Delta x_{dim(d_i)}}{2}}\cdots\int_{-\frac{\Delta x_1}{2}}^{\frac{\Delta x_1}{2}}\left(\varphi_j+\begin{bmatrix}u_1\\ \vdots \\ u_{dim(d_i)}\end{bmatrix}\right)du_1\cdots du_{dim(d_i)}\right\} \\ &= \frac{1}{q_i}\sum_{G_j \subseteq R_i}\left(\rho_j\varphi_j\right)\end{aligned} \qquad (12)$$

$$\hat{\Sigma}_i = \frac{1}{q_i}\sum_{G_j \subseteq R_i}\left\{\frac{\rho_j}{\prod_k \Delta x_k}\int_{-\frac{\Delta x_{dim(d_i)}}{2}}^{\frac{\Delta x_{dim(d_i)}}{2}} \cdots \int_{-\frac{\Delta x_1}{2}}^{\frac{\Delta x_1}{2}}\left(\varphi_j + \begin{bmatrix} u_1 \\ \vdots \\ u_{dim(d_i)} \end{bmatrix} - c_i\right)\cdot\left(\varphi_j + \begin{bmatrix} u_1 \\ \vdots \\ u_{dim(d_i)} \end{bmatrix} - c_i\right)^T du_1 \cdots du_{dim(d_i)}\right\}$$

$$= \frac{1}{12}\begin{bmatrix} {\Delta x_1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & {\Delta x_{dim(d_i)}}^2 \end{bmatrix} + \frac{1}{q_i}\sum_{G_j \subseteq R_i}\left[\rho_j(\varphi_j - c_i)(\varphi_j - c_i)^T\right] \tag{13}$$

$$i_k^* = \underset{i, R_i \in Adj(G^*)}{\operatorname{argmax}}\left[\frac{q_i}{\sum_i q_i} p\left(d_k \middle| d_k \sim N\left(\hat{\mu}_i, \hat{\Sigma}_i\right)\right)\right] \tag{14}$$

Supplementary Discussion: Equation Choice for Poisson Statistics

A. Statistical modeling of droplet counts in each cluster

As noted, if CPDs for target A and B are defined as λ_1 and λ_2, the probability that a droplet contains k_1 copies of target A and k_2 copies of target B can be calculated as equation (15), since target A and B enter droplets independently.

$$p(k_1,k_2) = \frac{\lambda_1^{k_1}}{k_1!}exp(-\lambda_1) \cdot \frac{\lambda_2^{k_2}}{k_2!}exp(-\lambda_2) \tag{15}$$

For each cluster, the droplet counts can be estimated with equation (15). We assume N: total droplet count, N_0: count of dual negative droplets, N_1: count of positive A/negative B droplets, N_2: count of negative A/positive B droplets, N_+: count of dual positive droplets. These droplet counts can be estimated as equation (16).

$$\begin{cases} N_0 = p(k_1 = 0,k_2 = 0) = Nexp(-\lambda_1)exp(-\lambda_2) \\ N_1 = p(k_1 > 0,k_2 = 0) = N[1 - exp(-\lambda_1)]exp(-\lambda_2) \\ N_2 = p(k_1 = 0,k_2 > 0) = Nexp(-\lambda_1)[1 - exp(-\lambda_2)] \\ N_+ = p(k_1 > 0,k_2 > 0) = N[1 - exp(-\lambda_1)][1 - exp(-\lambda_2)] \end{cases} \tag{16}$$

B. Three equation sets for Poisson statistics

There are two unknown parameters λ_1 and λ_2 in equation (16). Therefore, using two or more of the equations can derive these parameter, which results in different equation sets when different equations are chosen.

The first equation set (17) can be derived by choosing the second and third equations in the equation set (16), which uses N, N_1 and N_2 for calculation.

$$\begin{cases}\lambda_1 =- ln\left\{\frac{1}{2}\left[1 + \frac{N_2 - N_1}{N} \pm \sqrt{\left(1 + \frac{N_2 - N_1}{N}\right)^2 - \frac{4N_2}{N}}\right]\right\} \\ \lambda_2 =- ln\left\{\frac{1}{2}\left[1 + \frac{N_1 - N_2}{N} \pm \sqrt{\left(1 + \frac{N_1 - N_2}{N}\right)^2 - \frac{4N_1}{N}}\right]\right\}\end{cases} \quad (17)$$

There are two considerations when using the equation set (17). First, it is required that the discriminant $\Delta = 1 - 2(N_1 + N_2)/N + (N_1 - N_2)^2/N^2 \geq 0$. Second, there are two cases in each equation in the set, and one of each is used in different cases. When $N_0/(N_0 + N_1) \geq [1 + (N_2 - N_1)/N]$, the " + " in the first equation is used; otherwise, the " − " sign counterpart is used. Similarly, when $N_0/(N_0 + N_2) \geq [1 + (N_1 - N_2)/N]$, the " + " in the second equation is used; otherwise, the " − " sign counterpart is used.

The second equation set (18) can be derived by choosing the first three equations in the equation set (16), which uses N_0, N_1 and N_2 for calculation.

$$\begin{cases}\lambda_1 = ln\left(1 + \frac{N_1}{N_0}\right) \\ \lambda_2 = ln\left(1 + \frac{N_2}{N_0}\right)\end{cases} \quad (18)$$

There is one consideration when using the equation set (18) that at least one negative droplet is detected.

The third equation set (19) can be derived by choosing the first three equations in the equation set (16), which uses N, N_0, N_1, N_2 or N, N_1, N_2, N_+ for calculation.

$$\begin{cases}\lambda_1 = ln\left(\frac{N}{N_0 + N_2}\right) =- ln\left(1 - \frac{N_1 + N_+}{N}\right) \\ \lambda_2 = ln\left(\frac{N}{N_0 + N_1}\right) =- ln\left(1 - \frac{N_2 + N_+}{N}\right)\end{cases} \quad (19)$$

There is one consideration when using the equation set (19) that at least one negative droplet is detected, or at least one "positive A/negative B" droplet and "negative A/positive B" droplet is detected.

C. Comparison and choice of the optimal equation set

By applying the three equation sets to the droplet counts listed in Table S1–2, the following tables were derived.

Sample-	N	Classification results				Mutant copy number			Wild type copy number		
replicate		N_0	N_1	N_2	N_+	Eq. (17)	Eq. (18)	Eq. (19)	Eq. (17)	Eq. (18)	Eq. (19)
a-1	15605	10118	0	5487	0	0.0	0.0	0.0	10194.7	10194.7	10194.7
a-2	18631	11869	0	6762	0	0.0	0.0	0.0	10609.3	10609.3	10609.3
a-3	17866	11549	0	6317	0	0.0	0.0	0.0	10265.9	10265.9	10265.9
a-4	18216	11691	0	6525	0	0.0	0.0	0.0	10434.8	10434.8	10434.8
b-1	17111	11054	2	6055	0	4.3	4.3	2.8	10278.8	10277.9	10276.4
b-2	17296	11223	4	6069	0	8.4	8.4	5.4	10172.9	10171.3	10168.3
b-3	11814	7441	0	4373	0	0.0	0.0	0.0	10877.2	10877.2	10877.2
b-4	14386	9433	0	4953	0	0.0	0.0	0.0	9930.4	9930.4	9930.4
c-1	13092	8368	10	4710	4	28.1	28.1	25.2	10507.9	10506.3	10503.3
c-2	12924	8343	3	4577	1	8.5	8.5	7.3	10291.3	10290.7	10289.5
c-3	11176	7147	4	4021	4	13.2	13.2	16.8	10500.5	10502.6	10506.3
c-4	14338	9101	9	5225	3	23.3	23.3	19.7	10677.1	10675.1	10671.5
d-1	16824	10849	33	5920	22	71.4	71.5	77.0	10243.0	10246.1	10251.7
d-2	12942	8192	25	4710	15	71.7	71.7	72.8	10687.0	10687.6	10688.8
d-3	15486	9949	36	5488	13	85.0	85.0	74.6	10342.1	10336.4	10325.9
d-4	11036	7022	18	3992	4	60.3	60.2	47.0	10598.6	10591.0	10577.8
e-1	12709	8095	229	4252	133	655.4	656.4	679.9	9920.4	9933.3	9956.9
e-2	16985	10614	303	5895	173	662.0	662.3	668.8	10389.9	10393.7	10400.2
e-3	11193	6846	191	4052	104	648.3	647.5	628.5	10950.9	10939.2	10920.2
e-4	18135	11335	310	6312	178	634.6	634.9	641.8	10411.7	10415.8	10422.7
f-1	11980	6128	1670	3260	922	5638.5	5670.6	5736.6	9984.8	10036.8	10102.9
f-2	16781	8491	2319	4716	1255	5704.8	5681.5	5635.3	10432.6	10393.9	10347.7
f-3	11735	6008	1663	3130	934	5683.4	5749.5	5885.5	9762.4	9867.1	10003.0
f-4	16989	8648	2291	4699	1351	5460.0	5529.6	5677.0	10093.8	10210.9	10358.4
g-1	15301	2267	7566	1249	4219	34986.9*	34524.4	34602.2	10540.9	10326.3	10404.1
g-2	14969	2198	7253	1302	4216	33584.9*	34319.4	34193.4	10593.1	10946.2	10820.3
g-3	13795	2076	6782	1118	3819	35788.5*	34138.3	34424.3	10906.0	10137.2	10423.2
g-4	16897	2472	8343	1447	4635	32445.3*	34727.2	34383.5	9778.0	10842.6	10498.9

* These results were calculated with the “ − ” sign counterpart in equation (17).

Sample-	N	Classification results				Mutant copy number			Wild type copy number		
replicate		N_0	N_1	N_2	N_+	Eq. (17)	Eq. (18)	Eq. (19)	Eq. (17)	Eq. (18)	Eq. (19)
a-1	14093	9610	0	4478	5	0.0	0.0	8.3**	8996.6	9000.5	9008.8
a-2	14392	9816	0	4575	1	0.0	0.0	1.6**	9001.3	9002.1	9003.7
a-3	15777	10683	0	5092	2	0.0	0.0	3.0**	9169.7	9171.1	9174.1

Sample-		Classification results				Mutant copy number			Wild type copy number		
replicate	N	N_0	N_1	N_2	N_+	Eq. (17)	Eq. (18)	Eq. (19)	Eq. (17)	Eq. (18)	Eq. (19)
a-4	15357	10360	0	4992	5	0.0	0.0	7.7**	9250.3	9254.0	9261.6
b-1	15902	11074	3	4823	2	6.4	6.4	7.4	8506.2	8506.6	8507.6
b-2	15266	10311	1	4946	8	2.3	2.3	13.9	9213.9	9219.5	9231.1
b-3	17425	12080	2	5339	4	3.9	3.9	8.1	8610.1	8612.0	8616.2
b-4	15749	11229	2	4518	0	4.2	4.2	3.0	7957.0	7956.5	7955.3
c-1	13893	9558	7	4320	8	17.2	17.2	25.4	8771.0	8774.7	8782.9
c-2	15276	10685	15	4572	4	33.0	33.0	29.3	8382.7	8381.1	8377.4
c-3	16785	11454	9	5317	5	18.5	18.5	19.6	8971.5	8972.1	8973.2
c-4	18213	12230	10	5966	7	19.2	19.2	22.0	9347.1	9348.5	9351.2
d-1	14966	10153	34	4765	14	78.7	78.7	75.6	9055.6	9054.1	9051.0
d-2	15273	10329	28	4901	15	63.7	63.7	66.3	9135.5	9136.7	9139.4
d-3	15006	10163	28	4795	20	64.7	64.7	75.4	9088.9	9093.9	9104.6
d-4	17345	11682	32	5618	13	64.4	64.4	61.1	9240.6	9239.0	9235.8
e-1	16804	11015	273	5348	168	574.2	576.1	625.7	9287.1	9312.1	9361.8
e-2	16132	10635	260	5104	133	567.9	568.3	580.3	9217.4	9223.3	9235.3
e-3	14401	9599	229	4458	115	554.2	554.7	568.9	8968.8	8975.6	8989.7
e-4	16409	10700	260	5305	144	564.1	564.9	586.6	9463.2	9474.3	9495.9
f-1	11244	5818	1897	2607	922	6557.7	6640.2	6791.3	8607.7	8711.7	8862.8
f-2	12753	6362	2222	3073	1096	7022.4	7048.4	7090.4	9239.8	9272.6	9314.5
f-3	12962	6579	2268	3017	1098	6906.9	6969.3	7074.7	8804.9	8881.5	8986.9
f-4	13410	6766	2324	3159	1161	6868.1	6947.4	7081.0	8916.5	9015.2	9148.8
g-1	16671	2760	8522	1307	4082	33724.9*	33128.9	33194.5	9377.2	9121.8	9187.3
g-2	14165	2349	7308	1087	3421	33882.8*	33263.4	33328.6	9209.6	8948.7	9013.9
g-3	15169	2535	7772	1196	3666	32990.6*	33003.1	33001.8	9088.4	9093.7	9092.5
g-4	13065	2182	6669	979	3235	36334.1*	32948.0	33389.4	10172.4	8721.1	9162.5

* These results were calculated with the “ − ” sign counterpart in equation (17).

** These results were false-positive due to non-specific amplification in *EGFR* T790M assays.

By comparison, equation set (18) is the optimal choice. The results derived from all equations were consistent when there were no mutants or non-specific amplification of wild types. When mutant copy number is low (e.g. 3–6000), the results derived from equation set (17) and (18) were more consistent. When mutant copy number is high (e.g. 10000–30000), the results derived from equation set (18) and (19) were more consistent. These results suggested that equation (18) might be the optimal choice among the three equation sets.

With other considerations, equation set (18) is still the optimal choice. On one hand, equation set (17) complicates the calculation since a decision has to be made towards the " ± " sign every time this equation is used, as denoted in asterisks in the tables above. To make things worse, the discriminant Δ is not always non-negative in some real cases. For example, in a two/four-ddPCR assay where $N = 14631, N_0 = 5076, N_1 = 4393, N_2 = 5061, N_+ = 101$ as previously reported [8], the discriminant $\Delta = - 0.29 < 0$, which results in no real solutions to equation set (17). Using equation set (18), we can derive that $\lambda_1 = 0.62$ and $\lambda_2 = 0.69$, both of which can be further used to calculate copy numbers. On the other hand, equation set (19) yields false-positive results when non-specific amplification is hardly evitable in T790M assays, as denoted in double-asterisks in the tables above. Consider a case where there is no T790M mutants: the non-amplification of T790M wild types yields dual positive data points in the two/four-ddPCR results, and these data points inevitably affect the calculation result of equation set (19). On the contrary, equation set (18) does not use any data counts related to dual positive clusters such as N or N_+, and therefore such false-positive results can be avoided. Although equation set (18) cannot be applied to ddPCR data without dual negative cluster, such case in practice is rarely seen. Therefore, equation set (18) is applicable to the calculation of copy number for almost any two/four-ddPCR data with good performance. In conclusion, equation set (18) is still the optimal choice among the three equation sets.

References

1. A. S. Whale, J. F. Huggett and S. Tzonev, *Biomol. Detect. Quantif.*, 2016, **10**, 15-23.
2. C. A. Milbury, Q. Zhong, J. Lin, M. Williams, J. Olson, D. R. Link and B. Hutchison, *Biomol. Detect. Quantif.*, 2014, **1**, 8-22.
3. H. A. Sturges, *J. Am. Stat. Assoc.*, 1926, **21**, 65-66.
4. S. Beucher and C. Lantuéj, Proceedings of the International Workshop on Image Processing, 1979.
5. Y. Tarabalka, J. Chanussot and J. A. Benediktsson, *Pattern Recogn.*, 2010, **43**, 2367-2379.
6. C. Rambabu, I. Chakrabarti and A. Mahanta, *IEEE Proceedings-Vision, Image and Signal Processing*, 2004, **151**, 224-234.
7. J. Fang, *Medical Statistics and Computer Experiments*, World Scientific, 2014.
8. C. H. Roberts, W. Jiang, J. Jayaraman, J. Trowsdale, M. J. Holland and J. A. Traherne, *Genome Med.*, 2014, **6**, 20.