Supplementary Material A

Contest Format

The contest involved four stages, as described below.

Stage 1: Release of the Training Data In first stage, a training dataset comprising 52 cases were released to the registrants on April 24, 2016 through a secure website portal[1]. The dataset consisted of IHC and H&E stained images and the ground truth (GT). The GT score and percentage cells with complete membrane staining for the released training dataset can be seen in Table 1. At this stage, most of the details regarding contest (like tasks, contest rules, contest forum details etc) were already posted to the contest website and the registered teams started their work on algorithms for Her2 scoring. The registration process remained open for five weeks. We also created a social-forum (Google group) for the participants to share their queries and to communicate with the organizers.

Stage 2: Release of Off-Site Test Data A dataset comprising 28 cases were selected for off-site testing. This test dataset was released on May 17, 2016 and consisted of IHC and H&E stained WSIs without the GT information to ensure a fair evaluation. Source code for performance assessment in both MATLAB and Python languages were also released to the registrants. The registrants were given more than a month after releasing the test data to finalize and submit their scoring methods for announced tasks.

Stage 3: Submission of Results (Off-Site) The deadline for submission of results for the test dataset was set to be June 21, 2016, a week before to the contest workshop. Each team had to submit results in a comma-separated values (CSV) file along with a maximum 2-page summary of their algorithms, a description of experimental setup, and some preliminary results. The participants were advised that the CSV file should contain the predicted Her2 score, the confidence value for predicted score and the percentage of cells with complete cell membrane

staining (PCMS) for each WSI in the test dataset. Each registrant was allowed to submit up to three sets of results. The submitted results were evaluated but results were not announced until the contest workshop was held.

Stage 4: Contest Workshop. The contest workshop was conducted in Nottingham in conjunction with the annual meeting of the Pathology Society of Great Britain and Ireland on June 29, 2016. The contest workshop covered three main events: a) a brief talk from the organizers and the participants where 6 teams were invited for a small presentation to give an overview about their approaches and experiments, b) announcement of the comparative results of algorithms for both off-site, and c) announcement of results for the *Man vs Machine* comparison as a part of the on-site contests. The remaining 6 cases (of the 86) were used for an on-site competition (although they were released one day before the contest workshop due to the computational requirements of some of the automated algorithms and their results are not discussed here). The complete tables of results are available on the contest website[1].

Case	Ground truth	FISH Results	Percentage cells with complete membrane staining irrespective of intensity
1	0	N/A	0%
4	2	Negative	60%
6	2	Positive	40%
9	3	N/A	70%
11	3	N/A	90%
12	1	N/A	5%
13	0	N/A	0%

[1] http://www.warwick.ac.uk/TIAlab/Her2Contest/

14	1	N/A	1%
15	1	N/A	2%
16	1	N/A	5%
18	0	N/A	0%
19	3	N/A	30%
22	3	N/A	90%
24	1	N/A	5%
25	2	Negative	30%
26	2	Positive	50%
27	3	N/A	80%
29	0	N/A	0%
30	3	N/A	90%
32	1	N/A	3%
33	3	N/A	100%
34	1	N/A	2%
35	3	N/A	90%
36	2	Positive	100%
38	3	N/A	90%
39	0	N/A	0%
40	2	Positive	60%
46	0	N/A	0%
47	1	N/A	5%
48	2	Positive	20%
49	2	Positive	30%

50	2	Positive	50%
52	0	N/A	0%
55	2	Positive	70%
57	0	N/A	0%
58	1	N/A	5%
61	3	N/A	90%
63	2	Borderline amplified	70%
65	1	N/A	2%
66	0	N/A	0%
67	2	Positive	30%
68	0	N/A	0%
70	0	N/A	0%
73	0	N/A	0%
74	2	Positive	10%
79	1	N/A	5%
82	3	N/A	80%
83	3	N/A	100%
84	3	N/A	70%
86	1	N/A	3%
87	0	N/A	0%
88	1	N/A	5%

Table 1: The ground truth score for 52 cases from the training dataset with percentage of cells with complete membrane staining. The boderline case 63 was deemed negative and the amplification ratio for Her2 over-expression was 1.92.