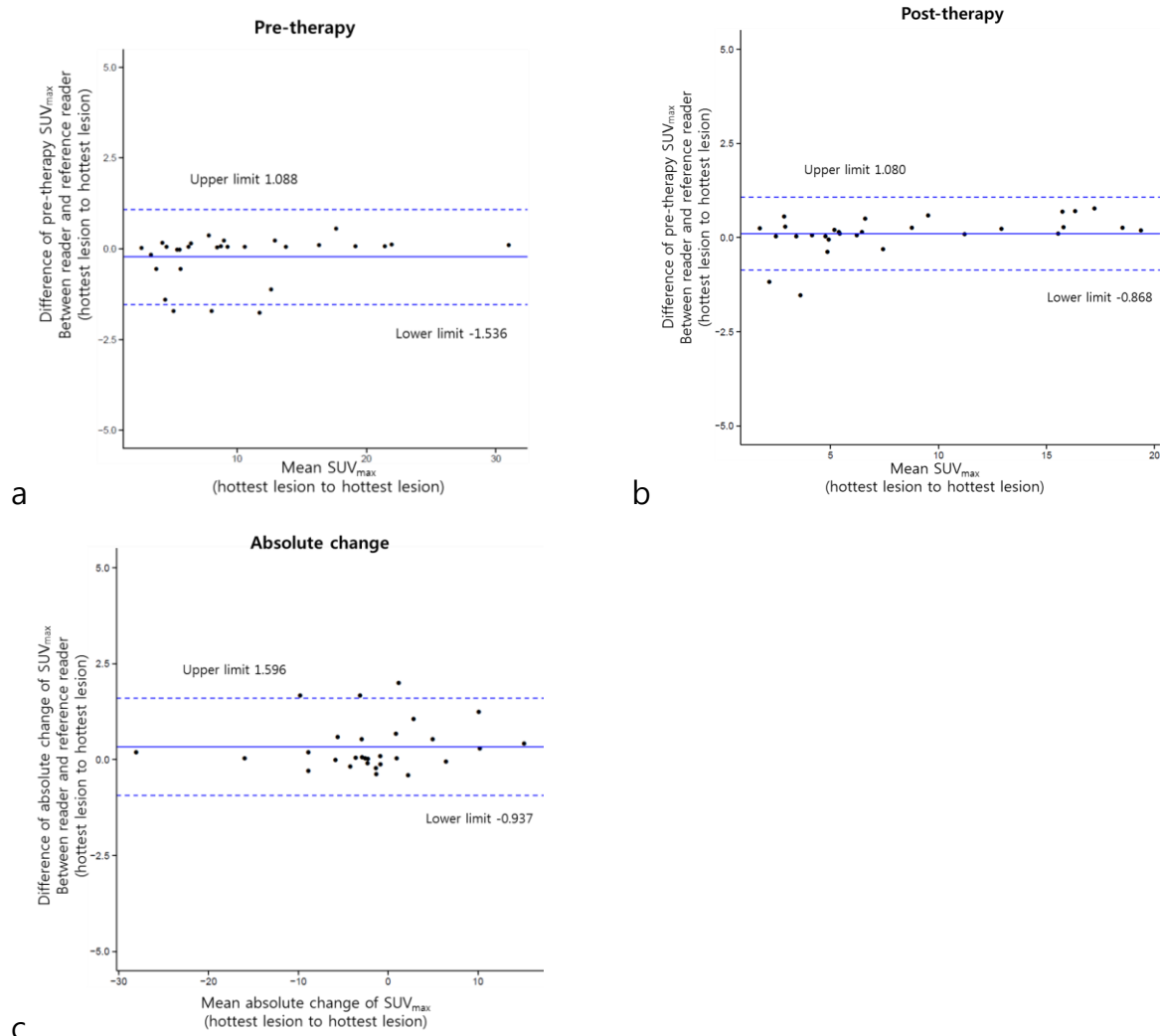
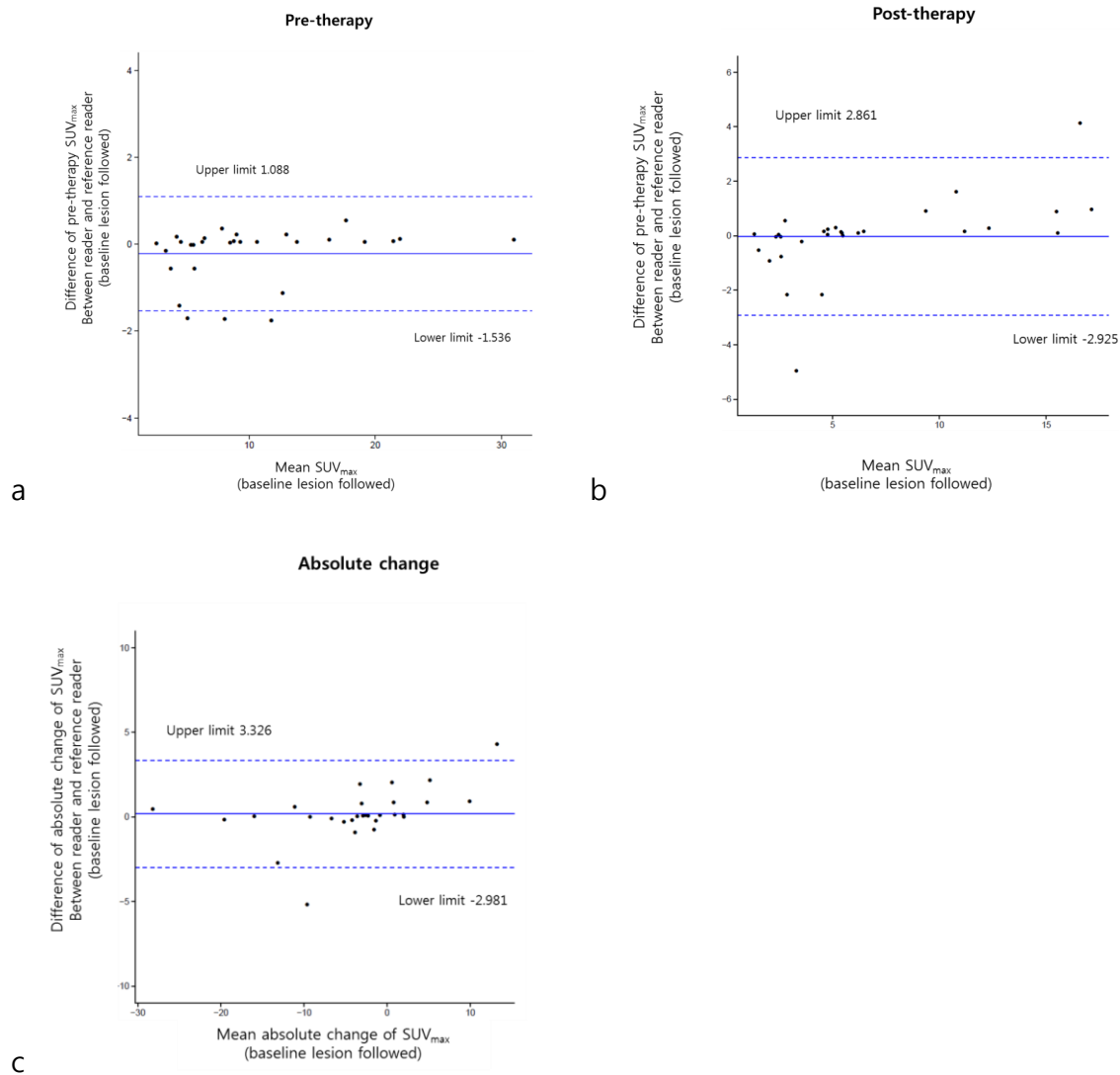


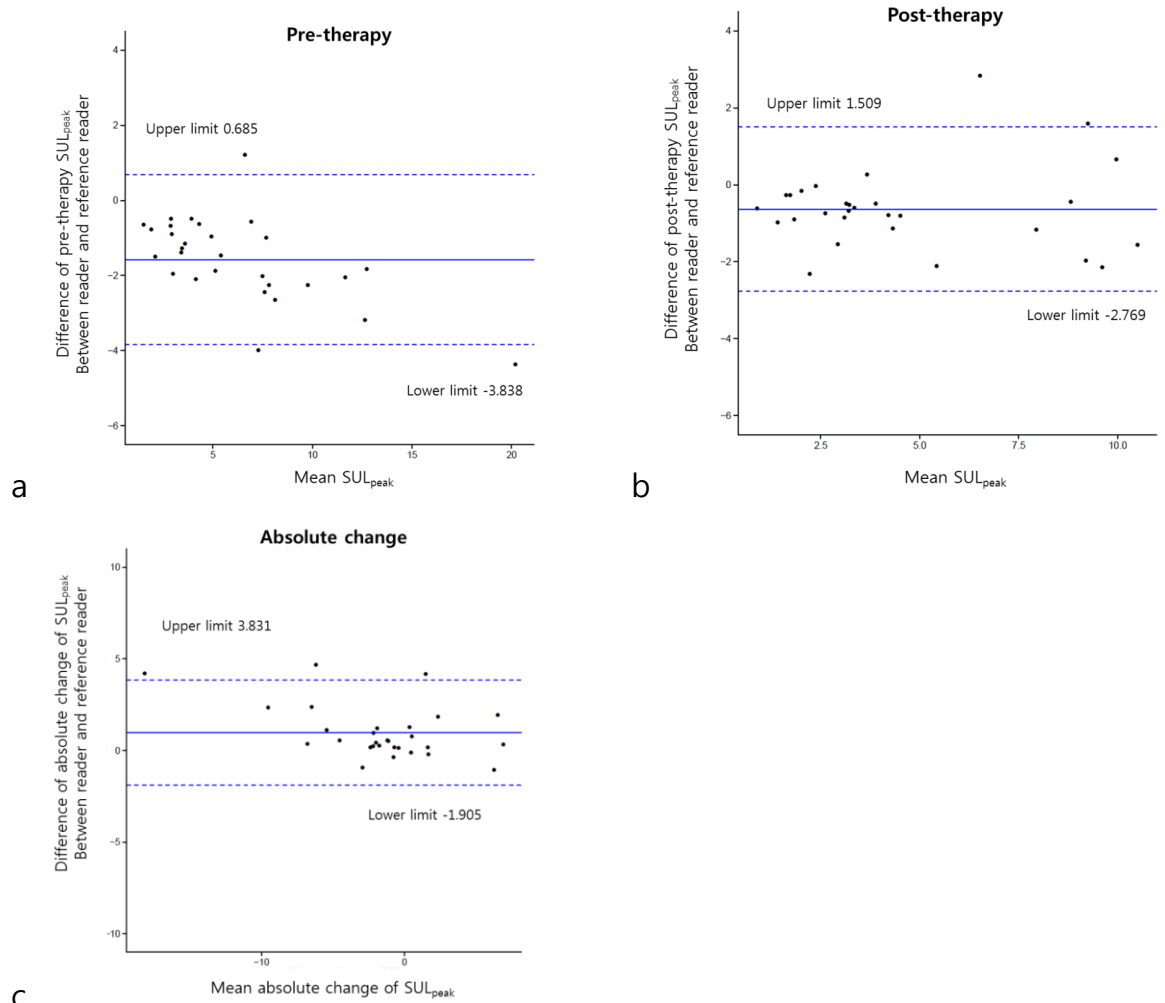
Supplemental figure 1. Bland-Altman plots of the  $SUV_{max}$  (*hottest lesion to hottest lesion*) measurements. The x-axis represents the average mean measurement by all readers. The y-axis represents the average difference between the 21 readers and the reference read. Each dot represents a case. The solid line represents the average bias, and the dashed lines represent the corresponding bias  $\pm 2$  standard deviations (2SD). (a)  $SUV_{max}$  at pre-treatment, (b)  $SUV_{max}$  at post-treatment, and (c) the absolute change of the  $SUV_{max}$  from pre- to post-treatment.



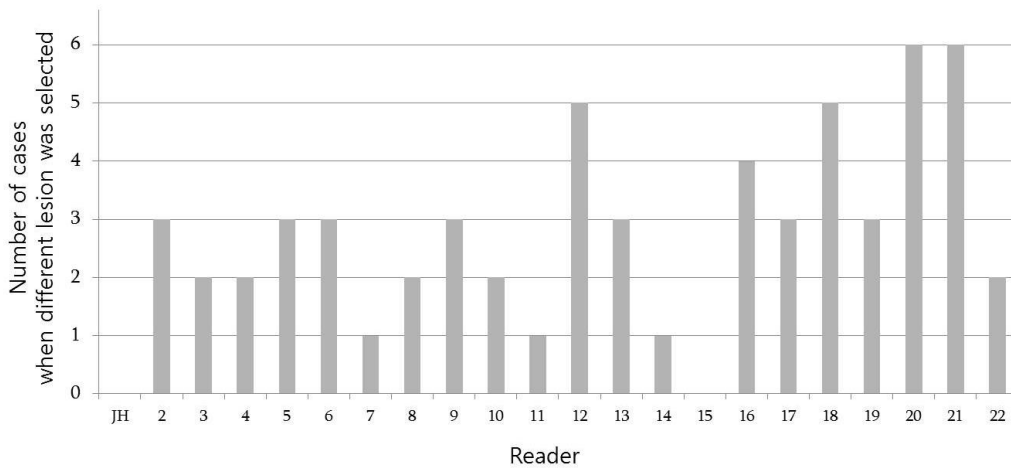
Supplemental figure 2. Bland-Altman plots of the  $SUV_{max}$  (*baseline lesion followed*) measurements. The x-axis represents the average mean measurement by all readers. The y-axis represents the average difference between the 21 readers and the reference read. Each dot represents a case. The solid line represents the average bias, and the dashed lines represent the corresponding bias  $\pm 2SD$ . (a)  $SUV_{max}$  at pre-treatment, (b)  $SUV_{max}$  at post-treatment, and (c) the absolute change of the  $SUV_{max}$  from pre- to post-treatment.



Supplemental figure 3. Bland-Altman plots of the  $SUL_{peak}$  (*hottest lesion to hottest lesion*) measurements. The x-axis represents the average mean measurement by all readers. The y-axis represents the average difference between the 21 readers and the reference read. Each dot represents a case. The solid line represents the average bias, and the dashed lines represent the corresponding bias  $\pm 2SD$ . (a)  $SUL_{peak}$  at pre-treatment, (b)  $SUL_{peak}$  at post-treatment, and (c) the absolute change of the  $SUL_{peak}$  from pre- to post-treatment.



Supplemental figure 4. The number of cases out of total 29 cases when a reader unaware of the medical history selected a different tumor lesion or benign condition as the “hottest tumor” compared to the reference read in the (a) pre-treatment PET/CT image, and (b) post-treatment PET/CT image. In the pre-treatment images, reader number 15 chose the same target lesion as the reference<sub>read</sub> in all cases.

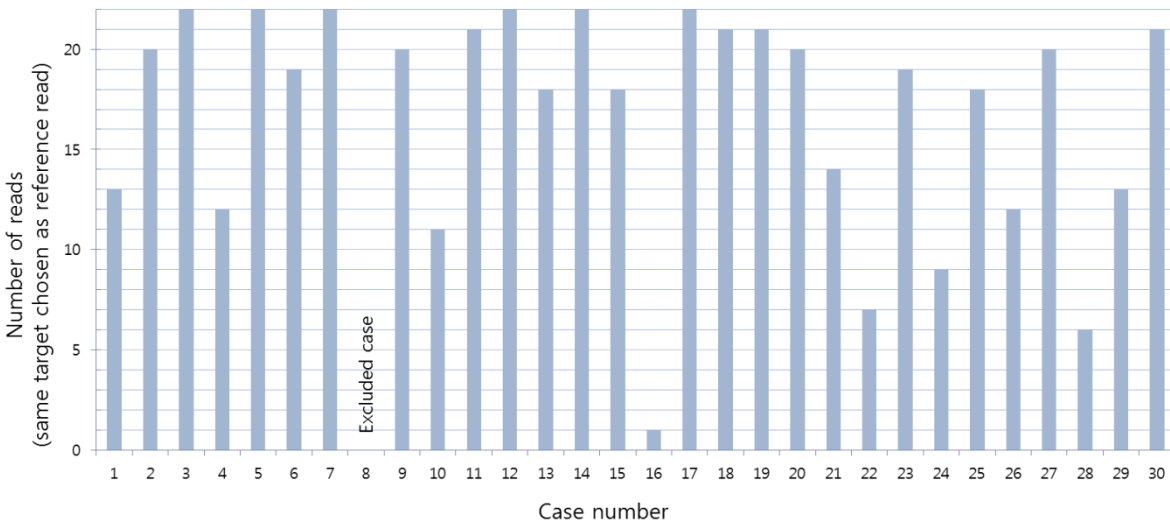


a

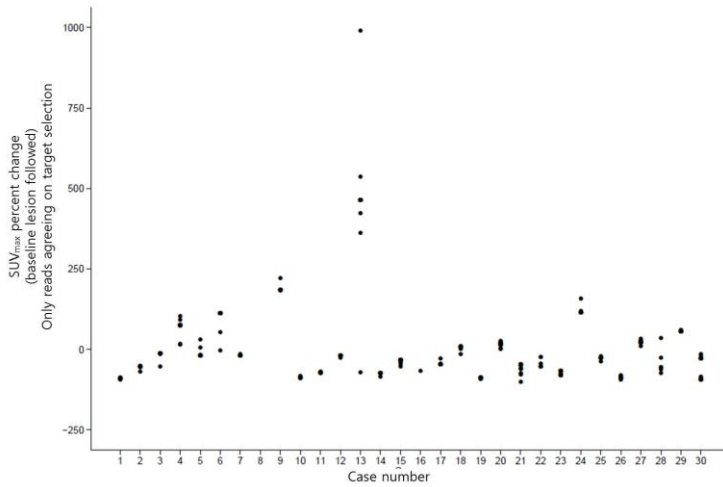


b

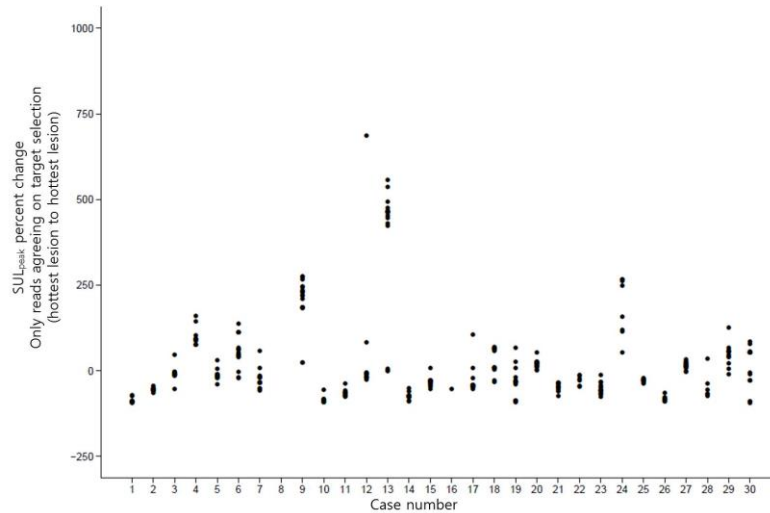
Supplemental figure 5. The number of *reads* per case with same target chosen as the reference *read*. There were total 22 possible *reads* per case. The readers all selected the same target in 6 cases. The cases with the highest number of *reads* disagreeing on the target selection had multiple tumor lesions. For example, cases number 16 and 28 had metastases in multiple bones, and the SUV difference among the lesions could be subtle. Cases number 10 and 22 also had extensive metastases involving multiple organs and nodal stations. There were multiple nodules throughout both lungs in case number 24. Case number 8 was excluded from analysis due to image corruption during the distribution process.



Supplemental figure 6. Data is plotted for only the cases with the same target lesion selected as the read<sub>reference</sub> (n=486). The percent change in the tumor FDG uptake from pre- to post-treatment was measured by (a) SUVmax (baseline lesion followed) and (b) SULpeak (hottest lesion to hottest lesion).

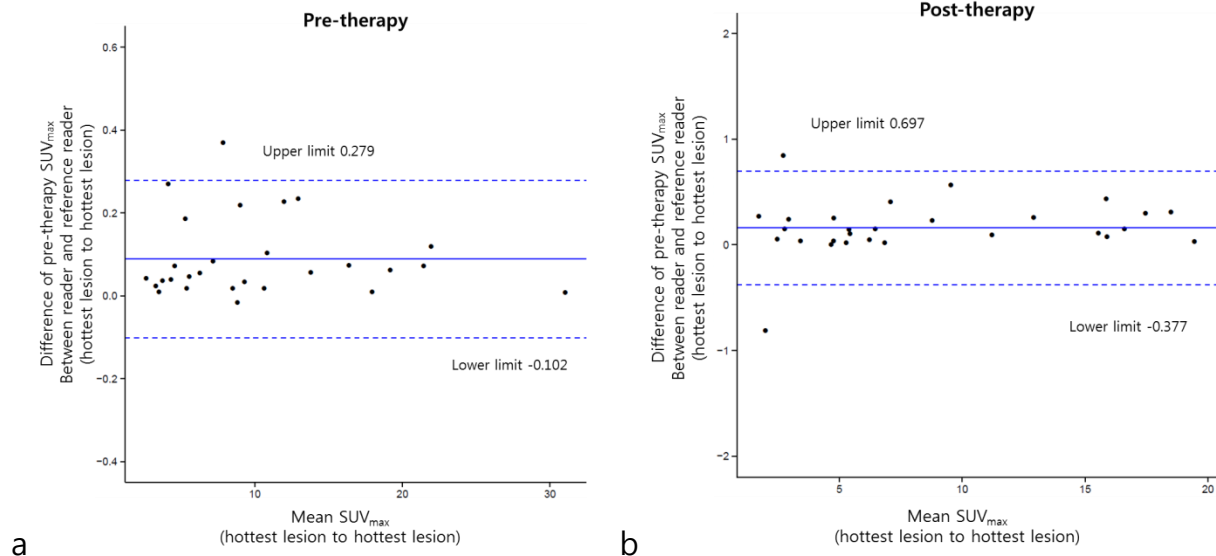


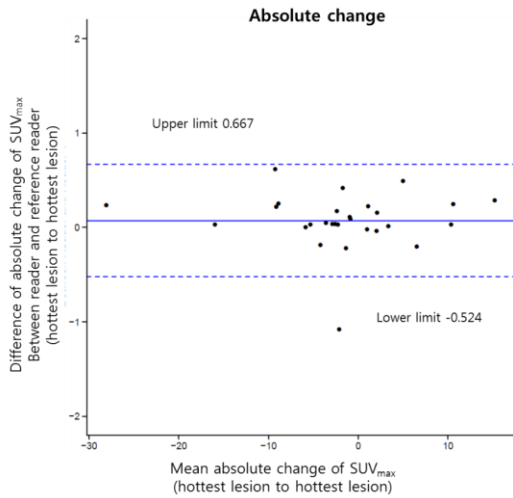
a.



b.

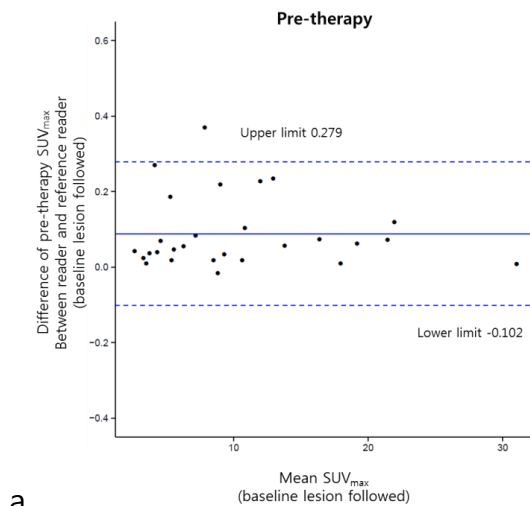
Supplemental figure 7. Bland-Altman plots of the  $SUV_{max}$  (*hottest lesion to hottest lesion*) measurements, with only the *reads* with the same target lesion selected as the reference read plotted. The x-axis represents the average mean measurement by all readers. The y-axis represents the average difference between the 21 readers and the reference read. Each dot represents a case (n=29). The solid line represents the average bias, and the dashed lines represent the corresponding bias  $\pm 2SD$ . (a)  $SUV_{max}$  at pre-treatment, (b)  $SUV_{max}$  at post-treatment, and (c) the absolute change of the  $SUV_{max}$  from pre- to post-treatment.



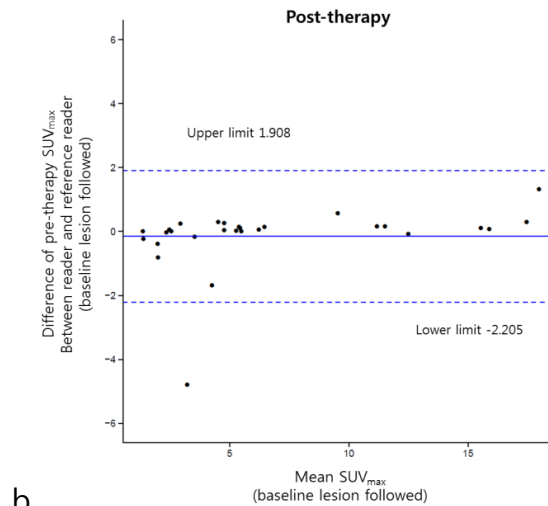


C

Supplemental figure 8. Bland-Altman plots of the  $SUV_{max}$  (*baseline lesion followed*) measurements, with only the *reads* with the same target lesion selected as the reference read plotted. The x-axis represents the average mean measurement by all readers. The y-axis represents the average difference between the 21 readers and the reference read. Each dot represents a case (n=29). The solid line represents the average bias, and the dashed lines represent the corresponding bias  $\pm 2SD$ . (a)  $SUV_{max}$  at pre-treatment, (b)  $SUV_{max}$  at post-treatment, and (c) the absolute change of the  $SUV_{max}$  from pre- to post-treatment.

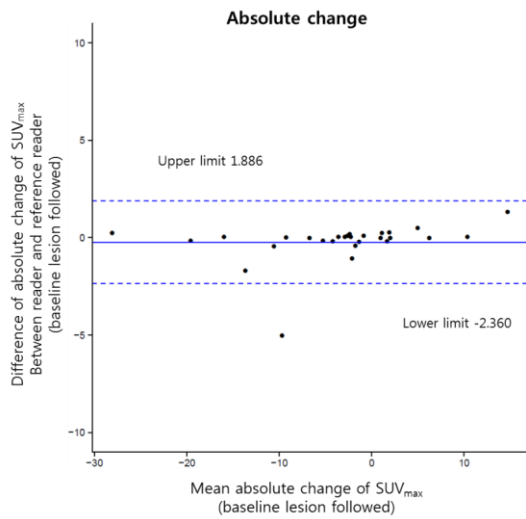


a



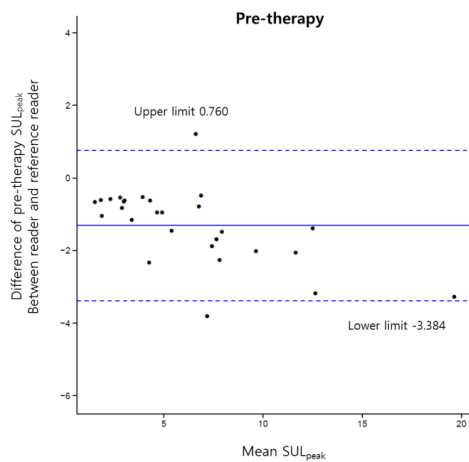
b



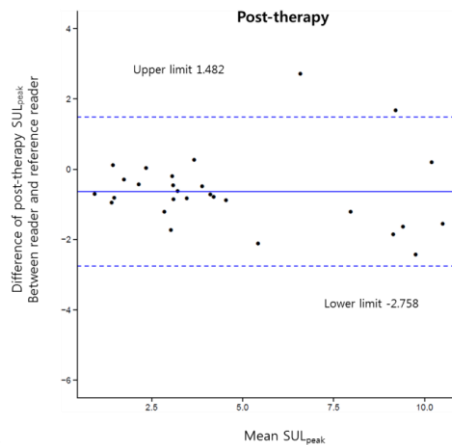


C

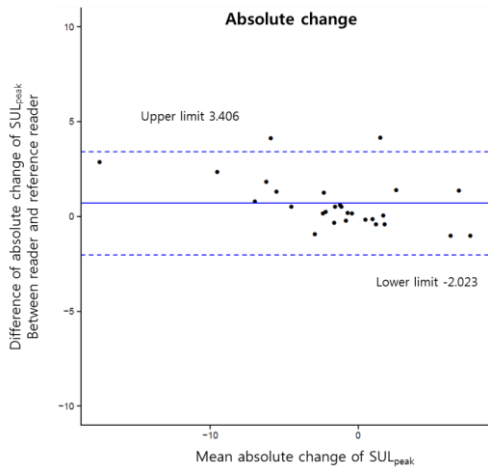
Supplemental figure 9. Bland-Altman plots of the  $SUV_{peak}$  measurements (*hottest lesion to hottest lesion*), with only the *reads* with the same target lesion selected as the reference read plotted. The x-axis represents the average mean measurement by all readers. The y-axis represents the average difference between the 21 readers and the reference read. Each dot represents a case. The solid line represents the average bias, and the dashed lines represent the corresponding bias  $\pm 2SD$ . (a)  $SUV_{peak}$  at pre-treatment, (b)  $SUV_{peak}$  at post-treatment, and (c) the absolute change of the  $SUV_{peak}$  from pre- to post-treatment.



a

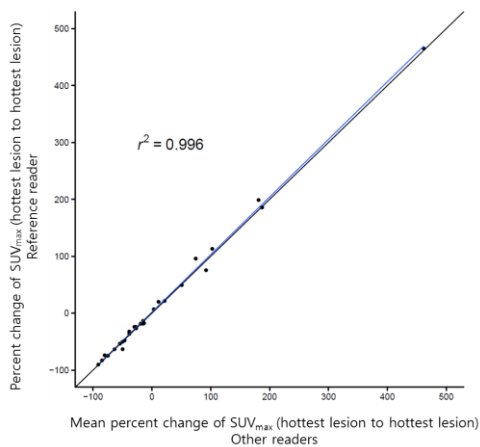


b

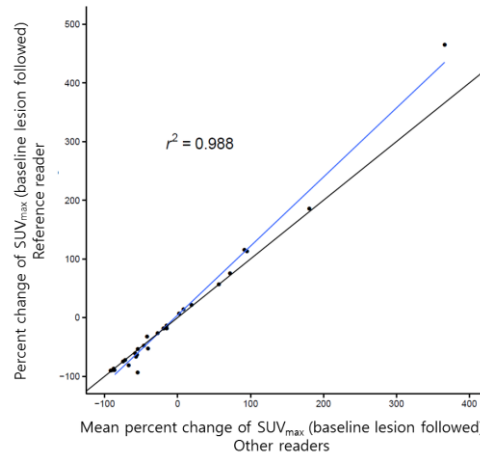


C

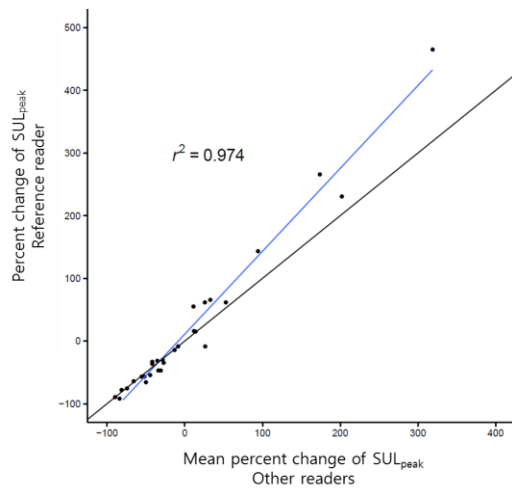
Supplemental figure 10. Graphical display of agreement between the reference<sub>read</sub> and the other readers. The reference<sub>read</sub> measurements were made while the reader was aware of the full medical history, while the other 21 readers were unaware. The mean percent change measurements of the 21 readers for each case, compared to the reference<sub>read</sub> for following PET parameters: (a)  $SUV_{max}$  (*hottest lesion to hottest lesion*), (b)  $SUV_{max}$  (*baseline lesion followed*), and (c)  $SUL_{peak}$  (*hottest lesion to hottest lesion*). A line from the linear model is included in blue, and the line of perfect agreement is in black.  $r^2$  is the correlation coefficient.



a

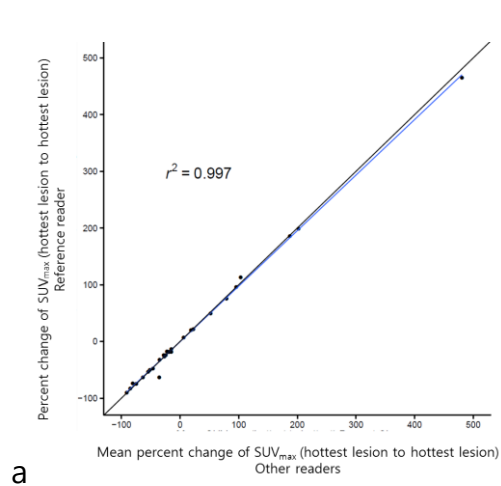


b

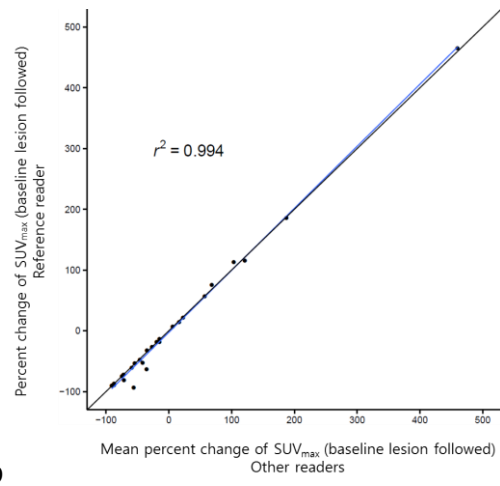


C

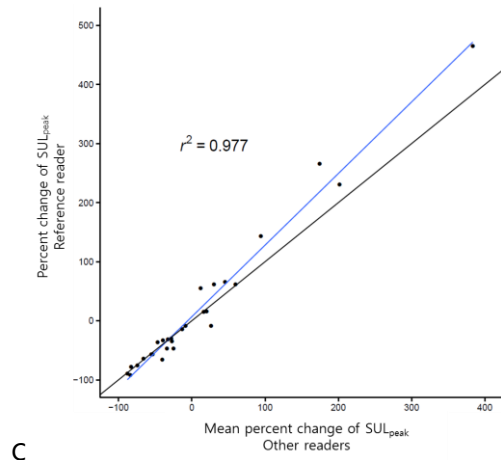
Supplemental figure 11. Graphical display of agreement between the reference<sub>read</sub> and the other readers, including only the *reads* with the same target lesion selected as the reference<sub>read</sub>. The reference<sub>read</sub> measurements were made while the reader was aware of the full medical history, while the other 21 readers were unaware. The mean percent change measurements of the 21 readers for each case, compared to the reference<sub>read</sub> for following PET parameters: (a)  $SUV_{max}$  (*hottest lesion to hottest lesion*), (b)  $SUV_{max}$  (*baseline lesion followed*), and (c)  $SUL_{peak}$  (*hottest lesion to hottest lesion*). A line from the linear model is included in blue, and the line of perfect agreement is in black.  $r^2$  is the correlation coefficient.



a



b



c