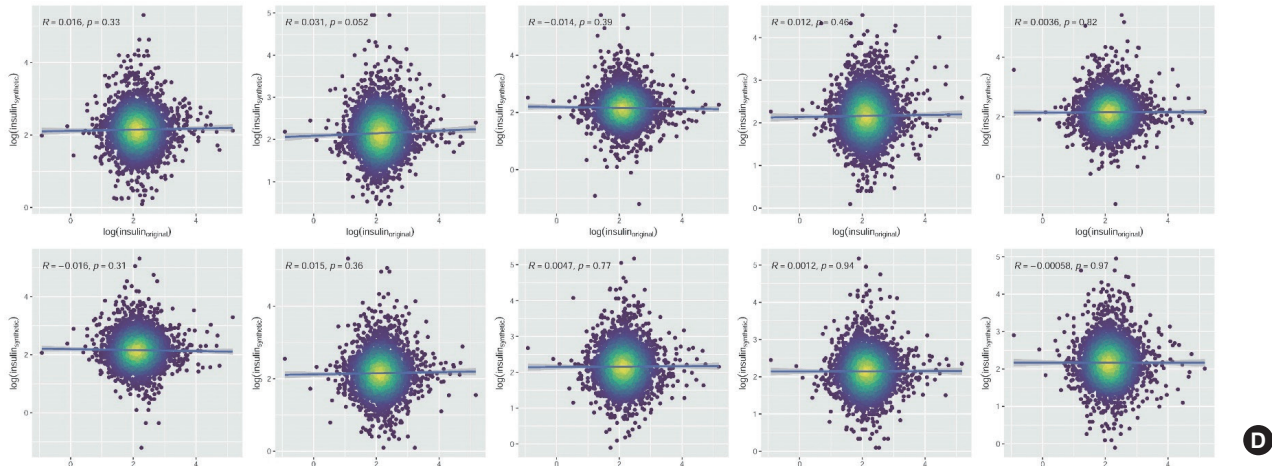**Supplemental Fig. S3.** *(Continued to the next page)*

Supplemental Fig. S3. (Continued) (A) Generation of validation dataset. The original Korea National Health and Nutrition Examination Survey (KNHANES) dataset (R, $n=15{,}629$), was split into a subset (Y) by randomly extracting 25% from the original dataset, and 10 synthetic datasets labeled R1, R2, ..., R10 were generated from the original data, respectively. (B) Statistical matching was performed between the subset of original dataset (Y) and the synthetic datasets (R1, R2, ..., R10) to obtain S1, S2, ..., S10, respectively. The scatterplot shows a weak but significant positive linear correlation (correlation coefficient $r=0.23$ to 0.25, $P<2.2\times10^{-16}$) between plasma insulin value from the original dataset (R) and that from the synthetic datasets (S1, S2, …, S10). When only (C) age- and sex-matching or (D) random matching was performed, no significant correlation was observed ($r=-0.019$ to 0.031, $P=0.052$ to 0.97).