Dataset generation for M&S

Qc Checklist

**1. Dataset generation strategy**

One file was provided (K24027-qPCR AAV DNA result (2-8주차).xlsx). To facilitate data tidying in R, xlsx file was converted to .csv format before importing them into R environment. The contents of the file were qPCR (Gene count) over time for various dose groups(G1~G11), with a total of 7 sites (in order: aqueous humor, vitreous, iris, retina, choroid, optic nerve, and serum) and concentration values divided into left and right eye. For ID, IDs specified in excel files were used as is. TIME was divided into time (TIME), day (DAY), and week (WEEK). For DV, concentration values that were BQL were all treated as 0 (DV=0, MDV=1). LEFT column (1: left, 0: right) was created to distinguish between left and right eye, for the SITE column, each site was represented numerically in the order mentioned above. For convenience, GROUP (G1 ~ G11), TYPE, BLQ (0), DRUG (4: Vehicle, 3: 3TF, 2: Aflibercept, 1: AIO), and ROW\_Number columns were added. After tidying, the dataset was arranged by ID, TIME, and DAY. The dataset generated using R was evaluated with the original dataset to identify the possible errors.

**2. Dataset generation**

Dataset generation was conducted using R (Version 4.3.1) and the `tidyverse` package (version 2.0.0). The final dataset consisted of 13 columns and 852 rows. The columns were named as specified above, and the dataset was arranged by ID, TIME, and DAY to create the final dataset. The R code used to generate the dataset is attached below as a .txt file (QC\_GC.txt, Supplementary 1).

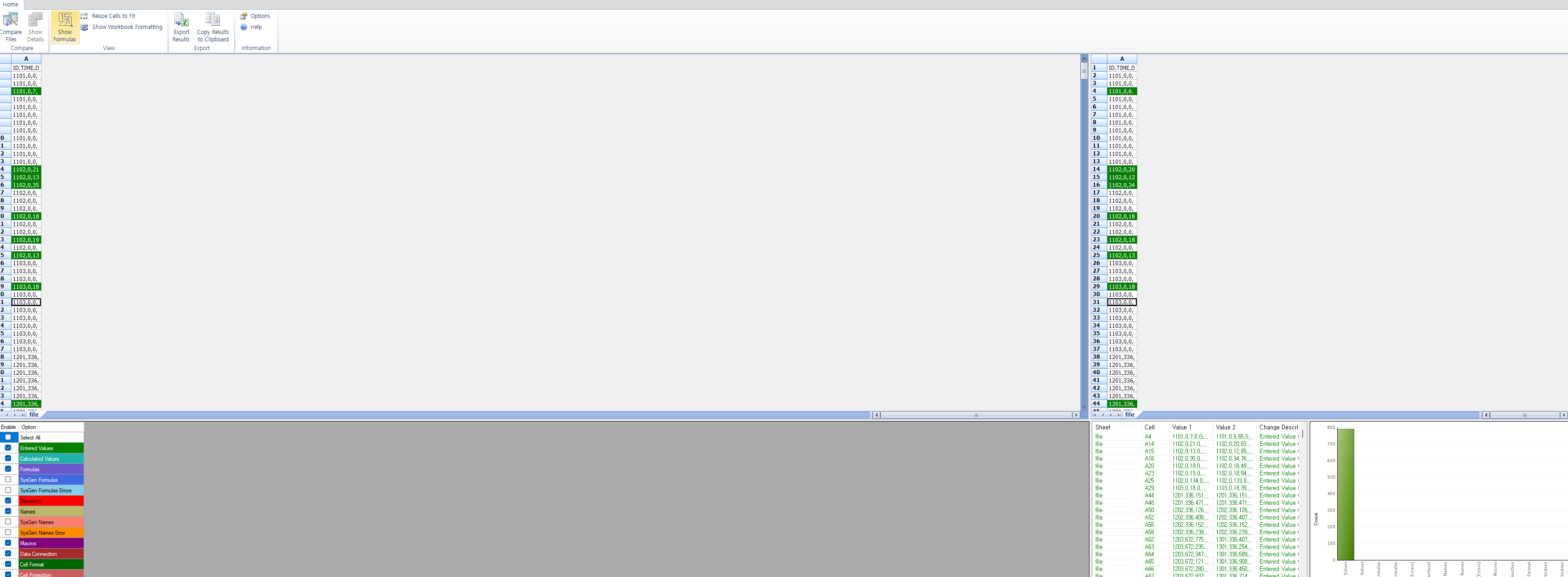
**2. Dataset evaluation**

Dataset comparison was performed using the `inquire` Addin provided in Excel (Figures). `Excel inquire` is an additional feature provided by Excel used for comparing different Excel files. Comparing two different datasets revealed that the number of rows and columns were all observed to be identical but `inquire` reported 292 unidentical rows (Inquire Report, Supplementary 2). This discrepancy appeared to be from variations in decimal point processing by R, causing nominally identical values to be recognized as different due to slight difference in decimal point. As the data generated by the above R code showed all decimal points as single digits, it appeared that errors related to decimal points occurred during the process of converting raw data to .csv and R tidying. No other notable differences were observed.

**2. Figures**

**텍스트, 스크린샷, 번호, 평행이(가) 표시된 사진

자동 생성된 설명**

****

Excel Addin `Inquire`