

Supplementary Figures

for manuscript

“Smoking induces DNA methylation changes in Multiple Sclerosis patients with exposure-response relationship”

Francesco Marabita, Malin Almgren, Louise K Sjöholm, Lara Kular, Yun Liu, Tojo James, Nimrod B Kiss, Andrew P Feinberg, Tomas Olsson, Ingrid Kockum, Lars Alfredsson, Tomas J Ekström, Maja Jagodic

Supplementary figure 1. Technical validation with bisulfite pyrosequencing. Three CpGs were validated with bisulfite pyrosequencing on DNA samples of the patients in the S cohort. All the CpGs in the PCR amplicon are shown and an asterisk indicates the CpG corresponding to the array probe, i.e. cg05575921 (AHRR), cg21566642 (ALPPL2) and cg06126421 (IER3). All differences in DNA methylation fractions between W5Y and NS are significant ($p < 0.001$, unequal variances T test).

Supplementary figure 2. Estimation of cell type proportions with a reference-based method for the S and B cohort. We estimated the cell type composition with the Houseman reference-based algorithm. The proportions of granulocytes, monocytes, B cells, natural killer cells, CD4 T cells and CD8 T cells are shown as boxplots (A-B). The scatterplots (C-D) show the P-values for the indicated contrasts with the inclusion of the estimated proportions in the association model (y axis) or as presented in the main text with the adjustment for the SVs (x axis). The points are colored using the local density and the Pearson correlation is shown on the bottom right of the plot.

Supplementary figure 3. Comparison with HLA-adjusted model for the B cohort. The scatterplots show the P-values for the indicated contrasts. Values on the x axis correspond to the values from the Surrogate Variable-adjusted analyses as presented in the main text, while values on the y result from the explicit adjustment for the HLA ("risk" vs. "no risk"). "Risk" patients are carriers of the *HLA-DRB1*15:01* allele (DR15+/+ or DR15+/-) and the non-carriers of the *HLA-A*02* allele (A2-/-). The points are colored using the local density and the Pearson correlation is shown on the bottom right of the plot.

Supplementary figure 4. "Concordance at the top" plots. For the i top-ranked CpGs of each list for the S and B cohort, concordance is defined as the fraction of the i probes in common between the S and B lists. This concordance is plotted as a function of i , for the ranked CpG lists obtained from the contrasts W5Y vs. NS (red), W5Y vs. B5Y (blue) and B5Y vs. NS (black). The inset shows an enlargement for the first 200 probes.

Supplementary figure 5. The effect of smoking on DNA methylation for active smokers. DNA methylation is shown for seven selected and replicated loci affected by smoking, separately for the S and B cohorts. One CpG per gene locus is shown, Methylation levels (β values) are shown in relation to the NS and W5Y smoking categories. The W5Y category was further stratified in active and non-active smokers, considering individual that reported smoking < 1 year from the time of sampling as active smokers.

Supplementary figure 6. Functional annotation of DMPs, with respect to DHSs. The enrichment – $\log_{10}$ P-values for the DHSs of the blood cell types from Encode (A) or NIH Roadmap Epigenomics (B) data was obtained with eFORGE, with default setting. Dots are colored according to the FDR and significant enrichments are labeled.

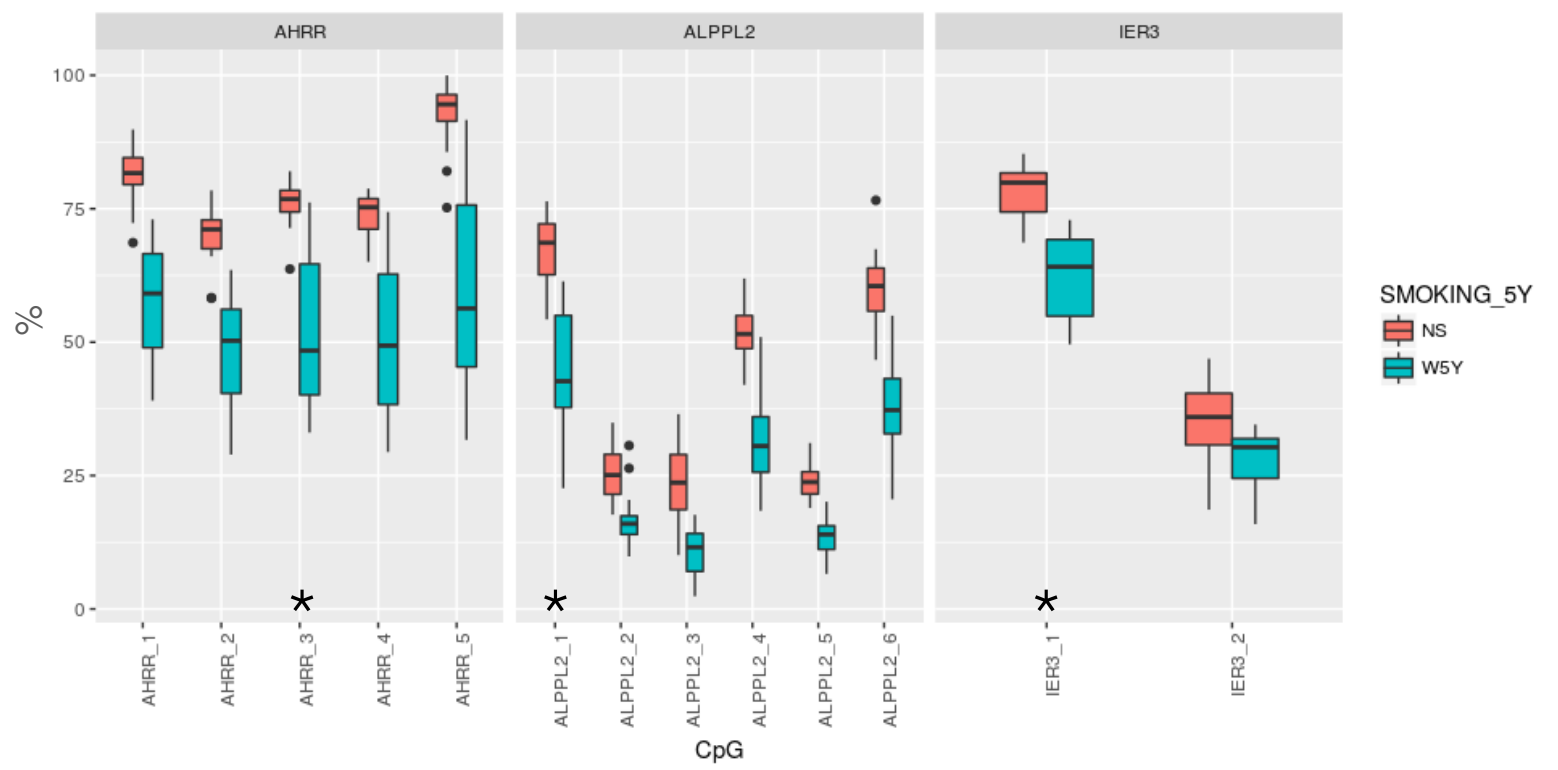
Supplementary figure 7. The 58 DMPs for the S cohort. The methylation levels (β values) for the 58 DMPs are shown in relation to the NS, B5Y and W5Y smoking categories or the time since smoking cessation.

Supplementary figure 8. The 58 DMPs for the B cohort. The methylation levels (β values) for the 58 DMPs are shown in relation to the NS, B5Y and W5Y smoking categories or the time since smoking cessation.

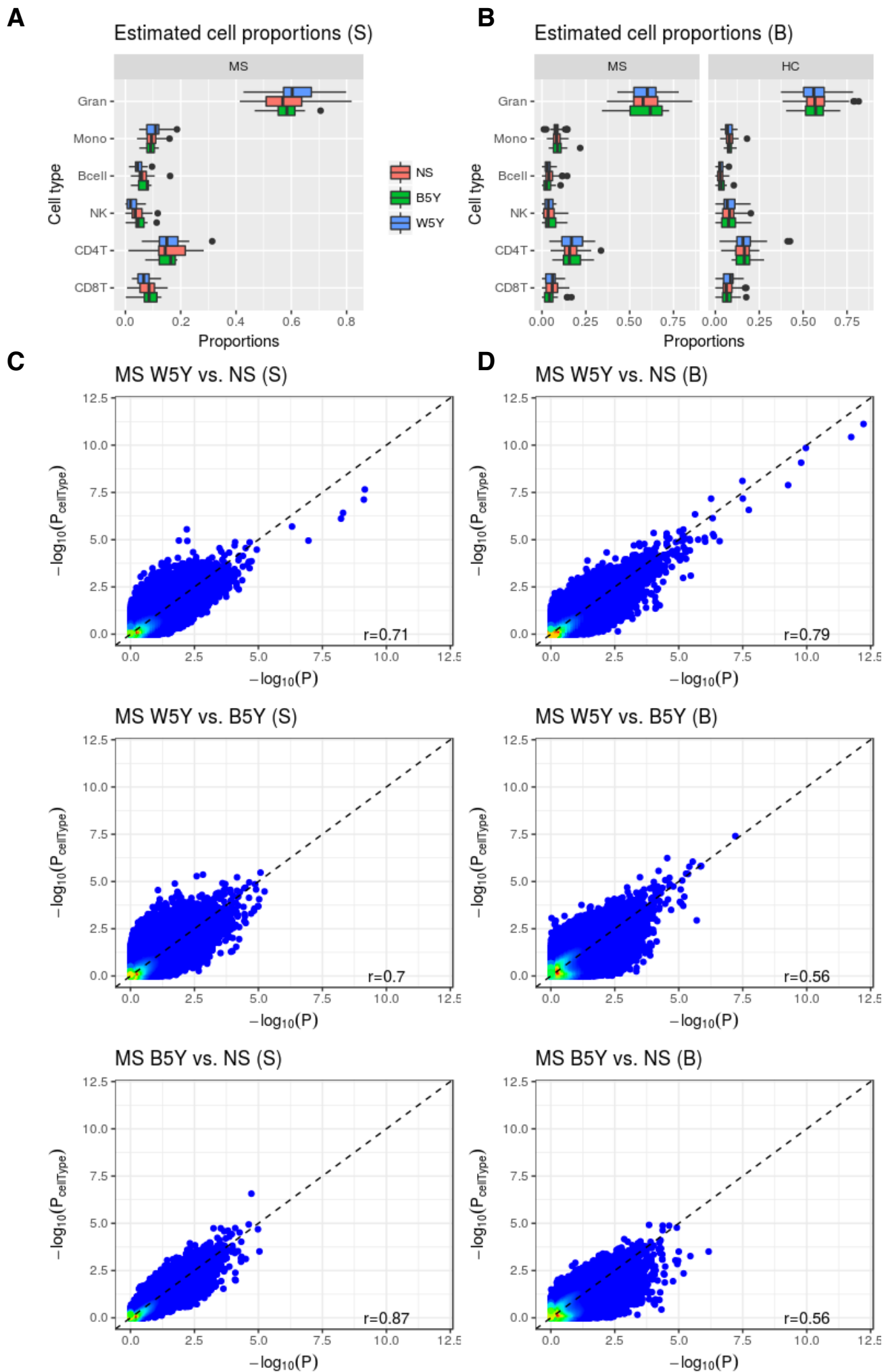
Supplementary figure 9. Gene expression of the AHRR gene in the PBMC cohort. *AHRR* gene expression levels are shown for the PBMC cohort as normalized counts. On the panels on the left, smoking is stratified as NS, B5Y and W5Y, while on the right smoking is stratified as follows: NS, never smokers; PS, patients that reported smoking >1 year from the time of sampling; CS, patients that reported smoking <1 year from the time of sampling. Grey dots mark individuals who reported smoking at the time of sampling. The PY-expression scatterplot is shown only for the patient in the W5Y or CS categories, respectively.

Supplementary figure 10. Enrichment of known smoking-affected CpGs for Healthy Controls samples. A GSEA approach was used to check the enrichment for known smoking-affected CpG. The green line on the top panels shows the Enrichment Score (ES). Vertical ticks mark the location of the 62 CpG from Gao *et al.* within the list of the probes ranked by decreasing significance in healthy controls and the bottom panels show their corresponding normalized $-\log_{10}P$ values, for the comparisons W5Y vs. NS, W5Y vs. B5Y and B5Y vs. NS, respectively.

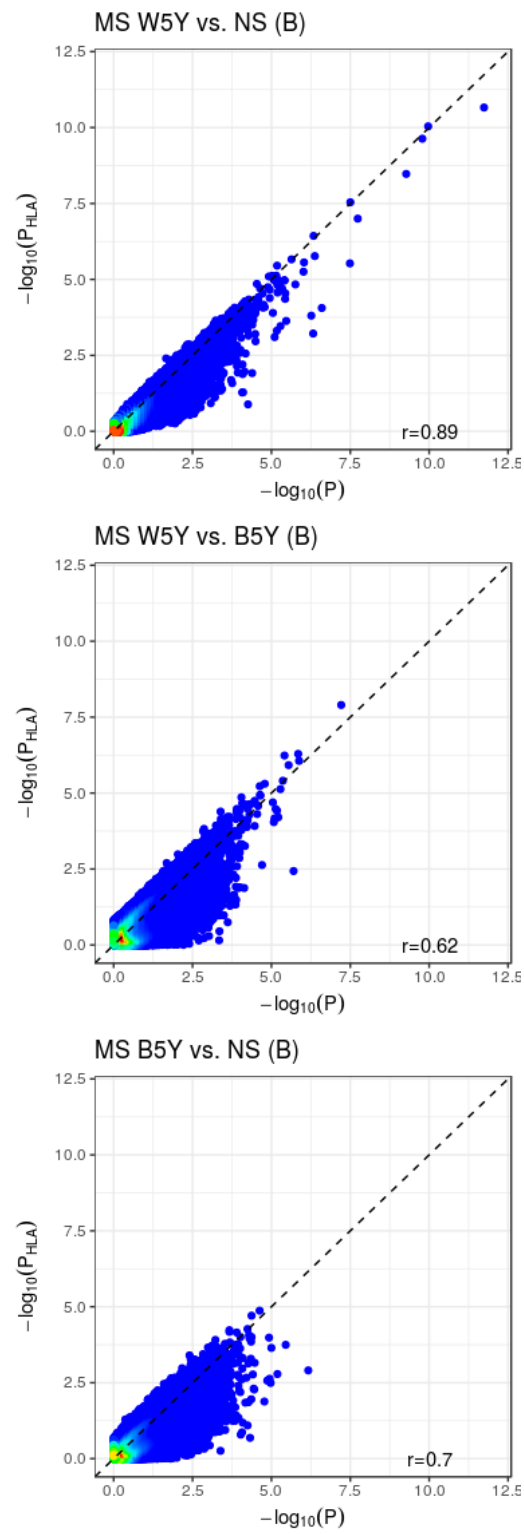
Supp. Fig. 1



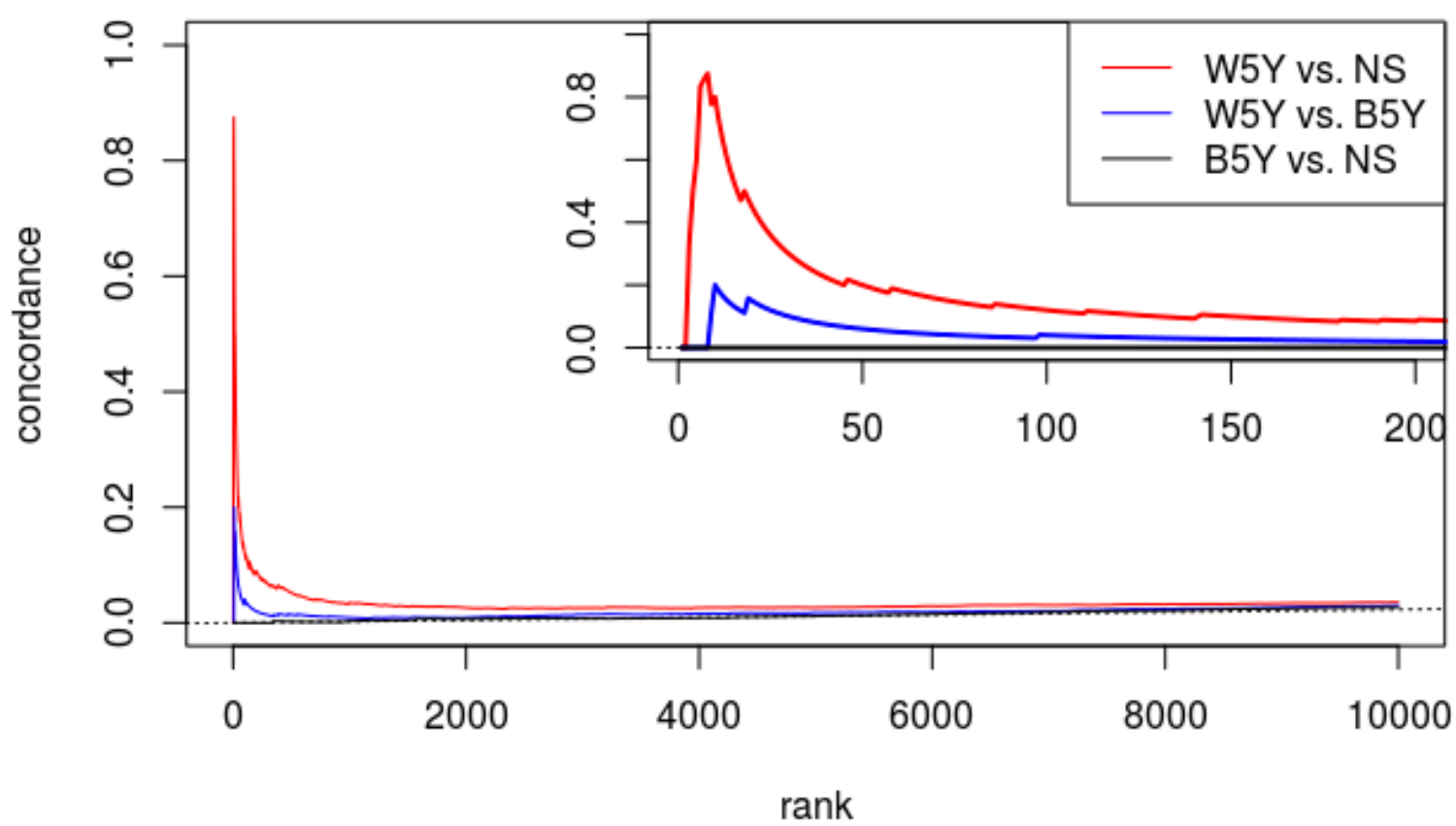
Supp. Fig. 2



Supp. Fig. 3

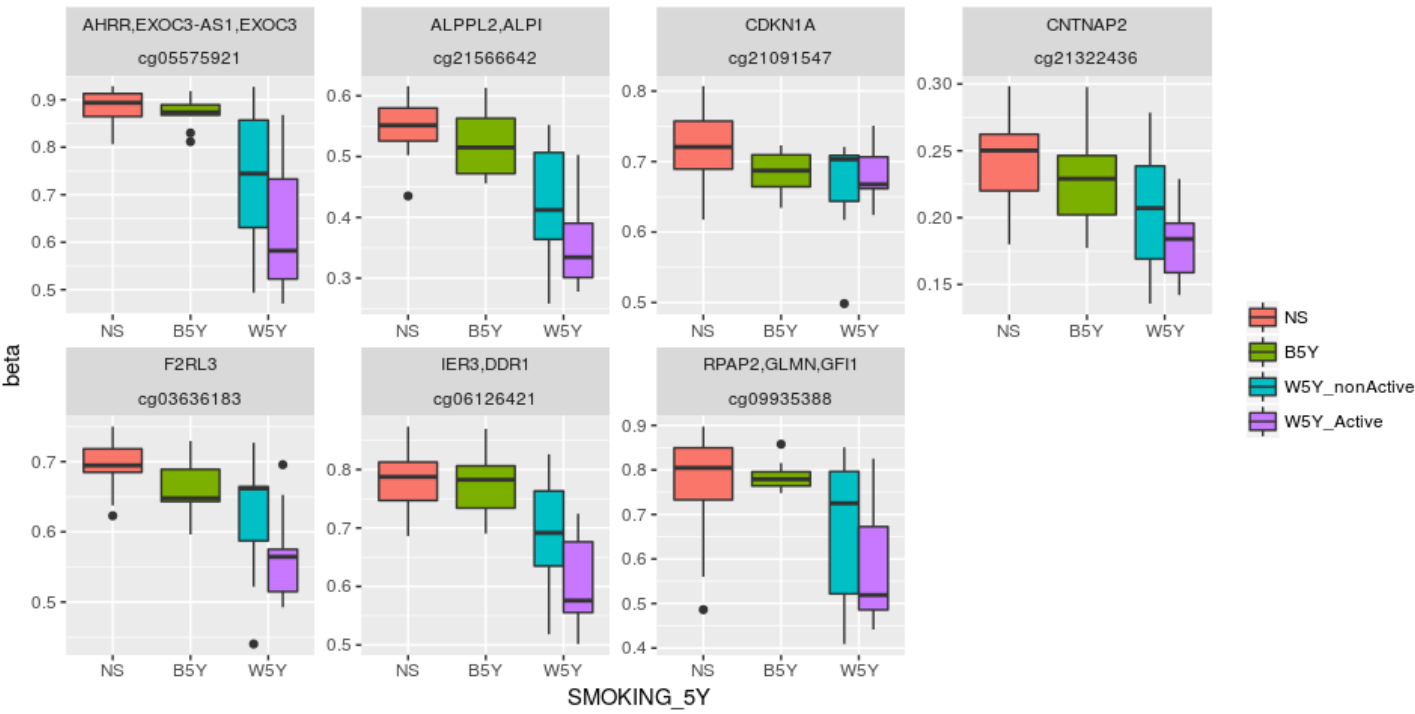


Supp. Fig. 4

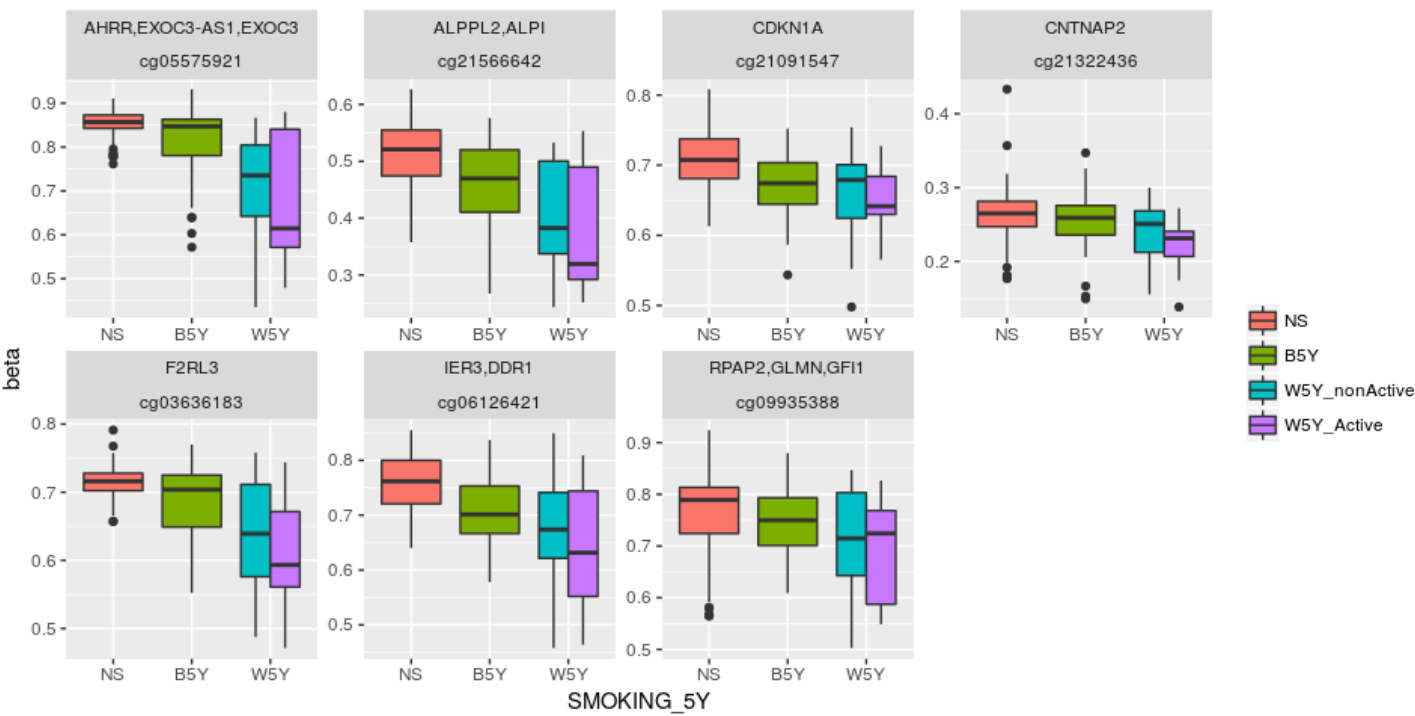


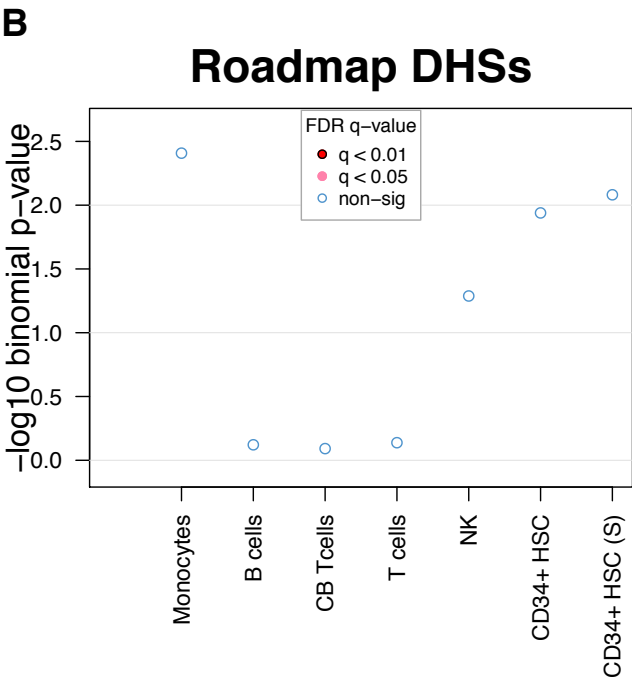
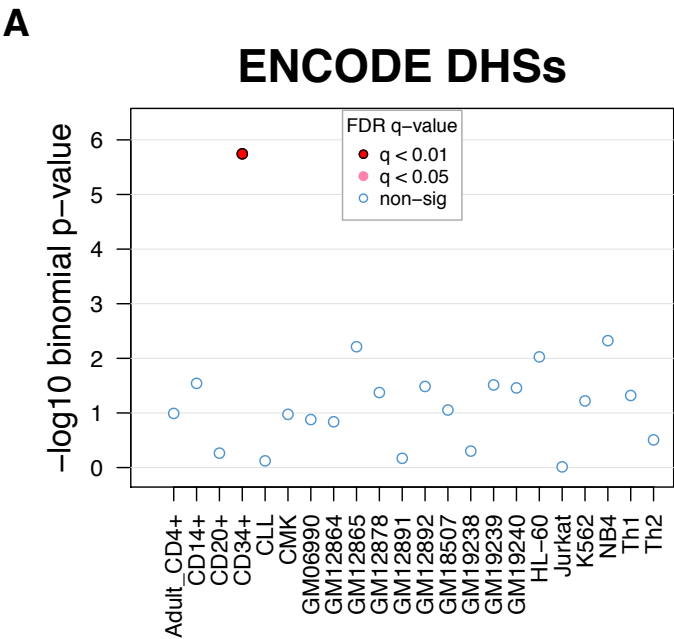
Supp. Fig. 5

S cohort

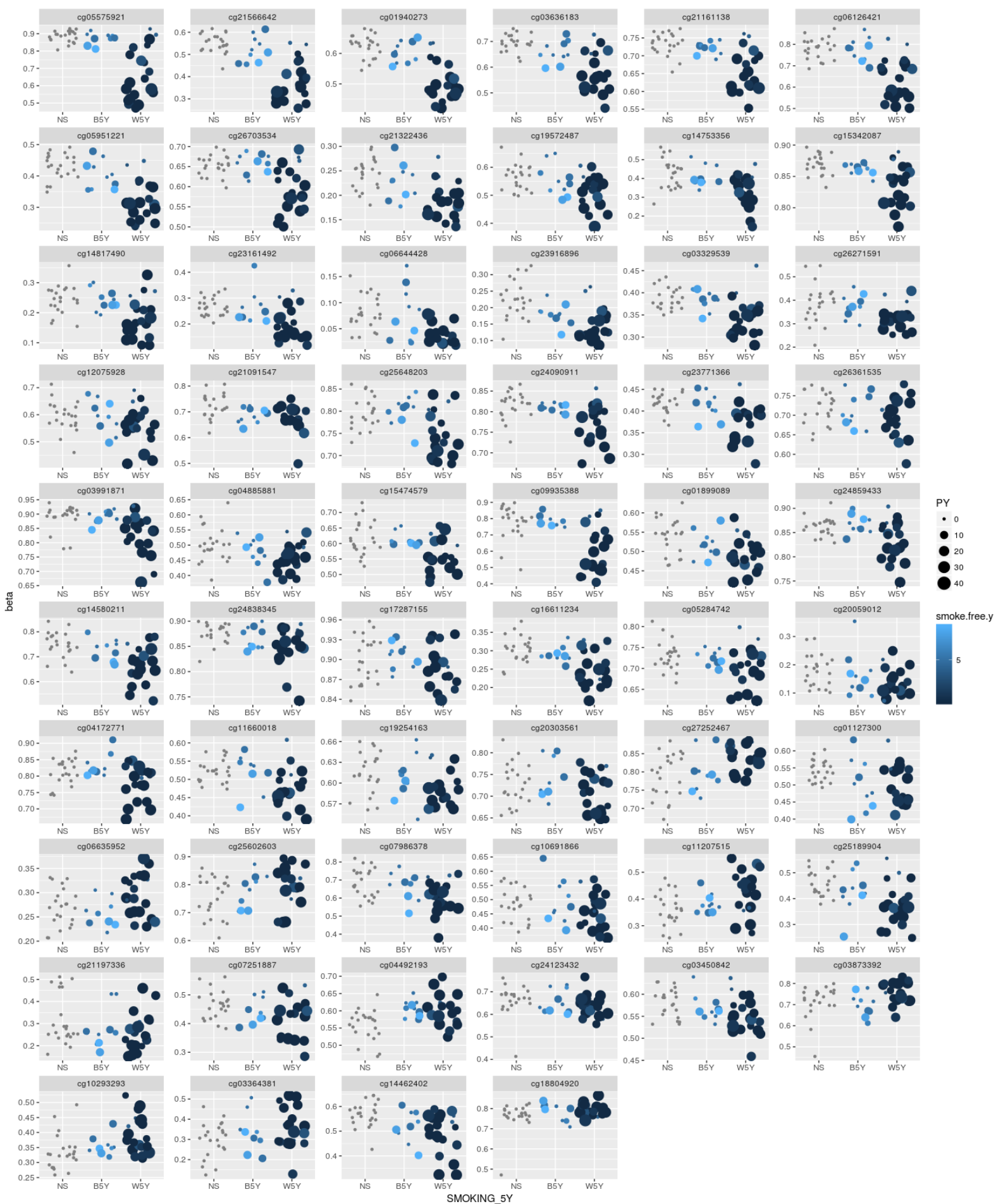


B cohort





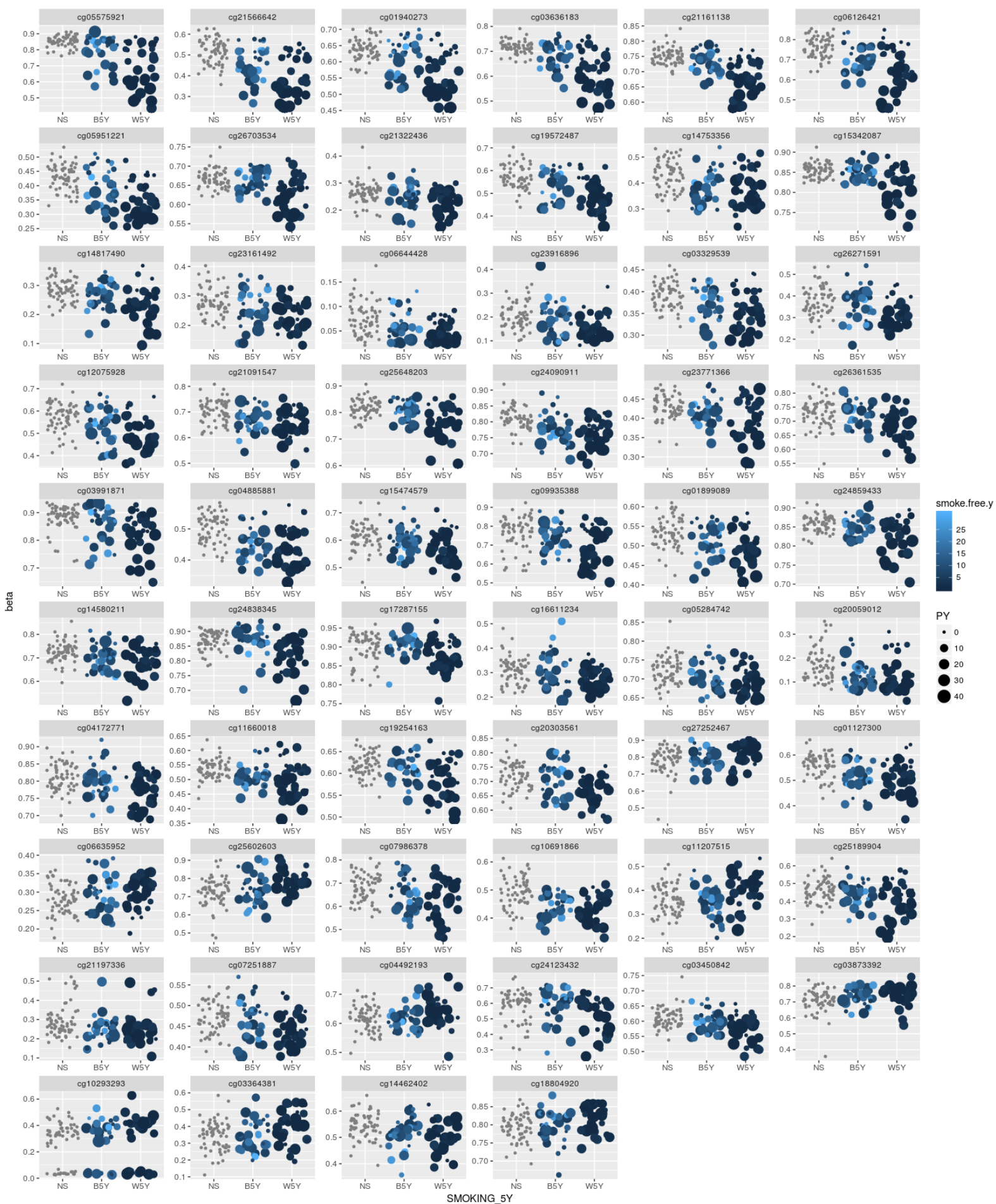
Supp. Fig 7



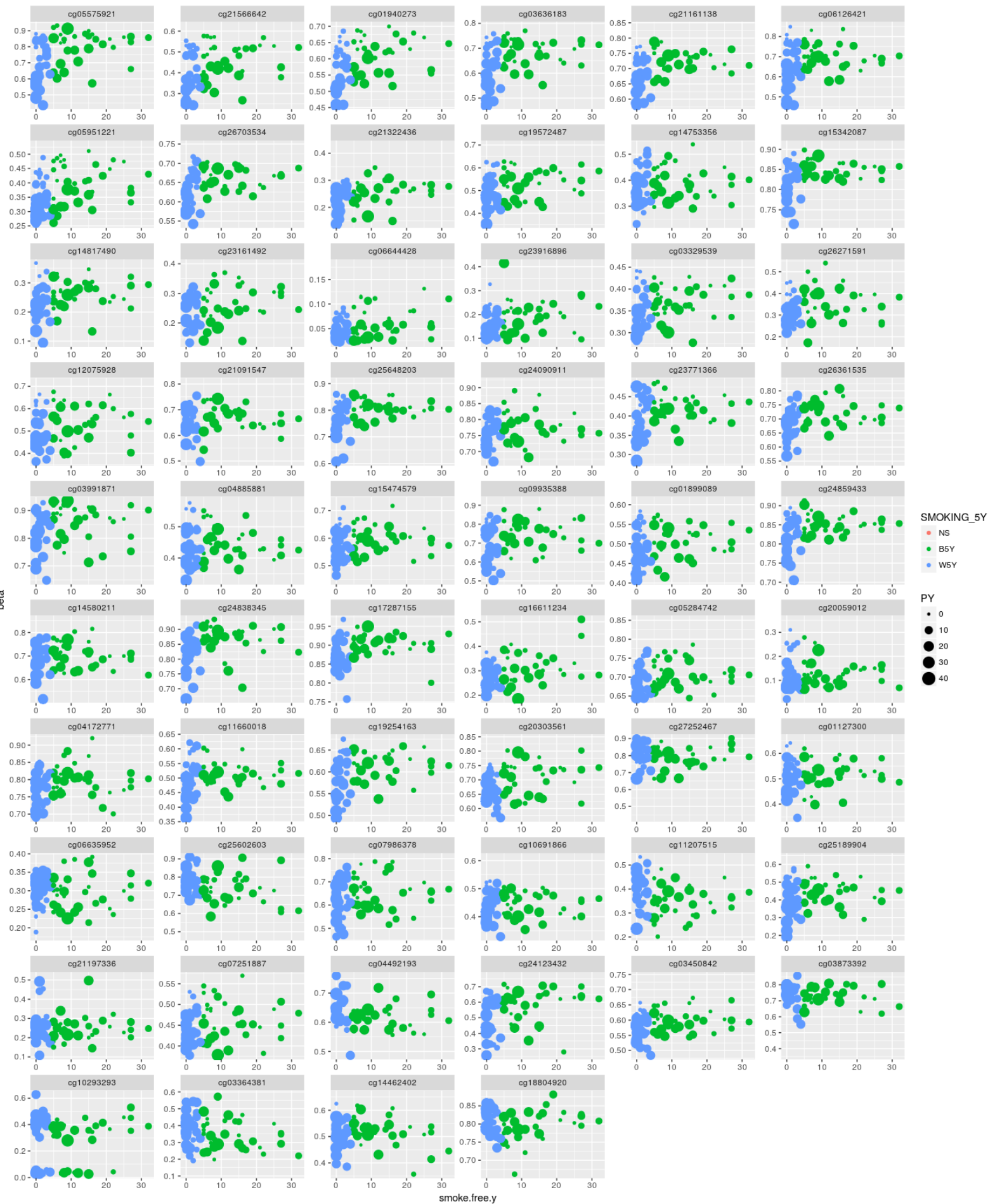
Supp. Fig 7 (continued)



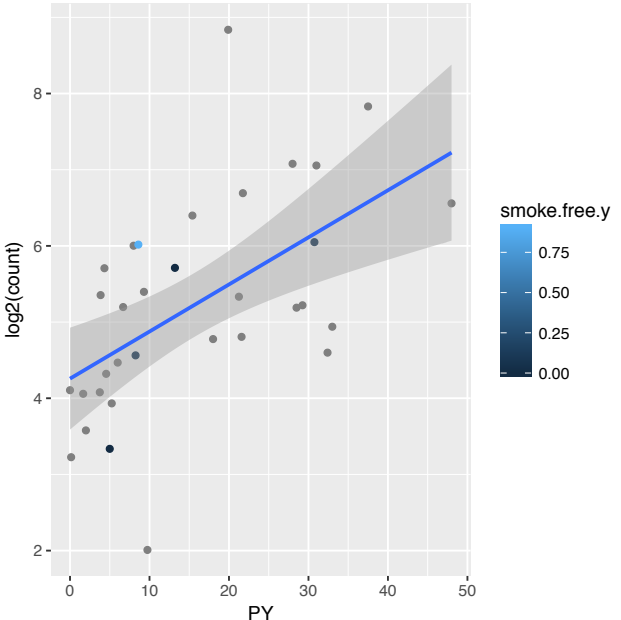
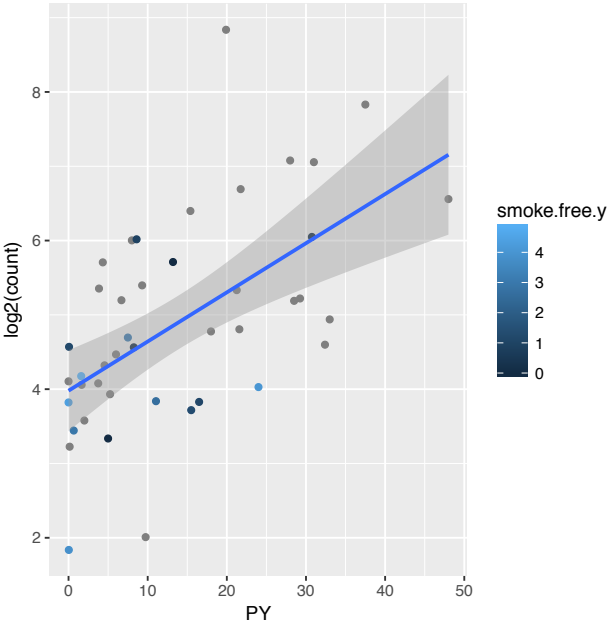
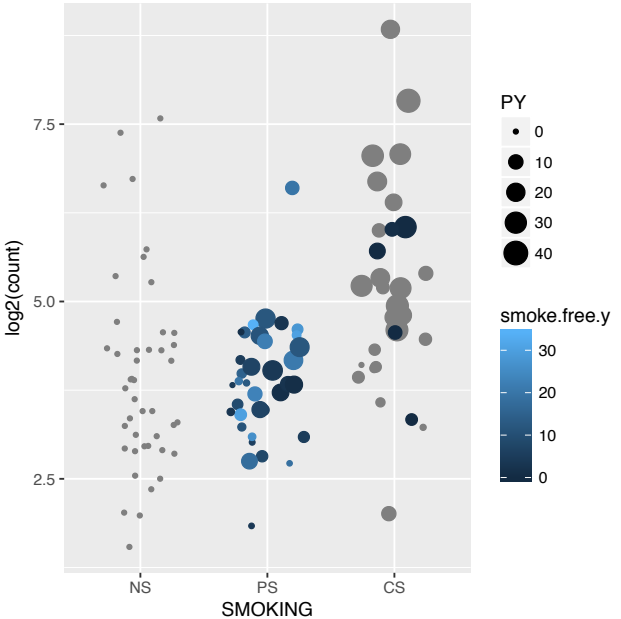
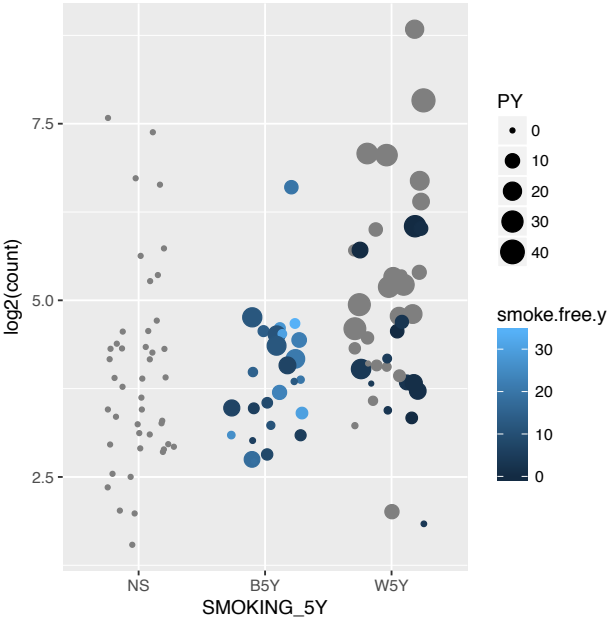
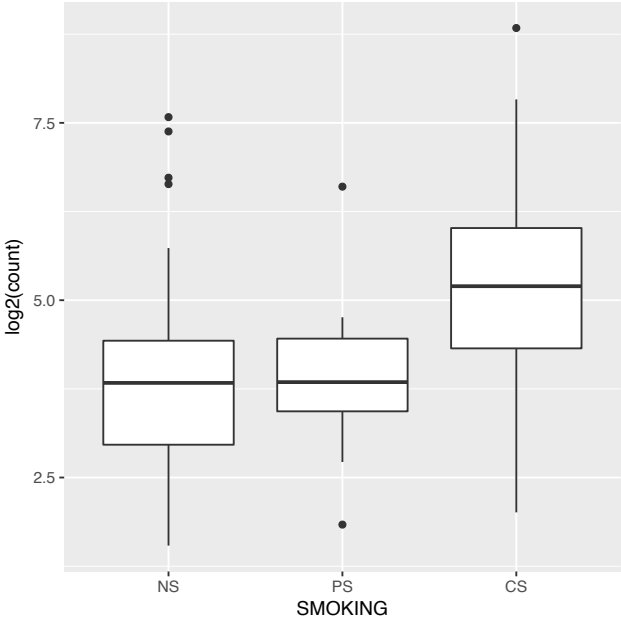
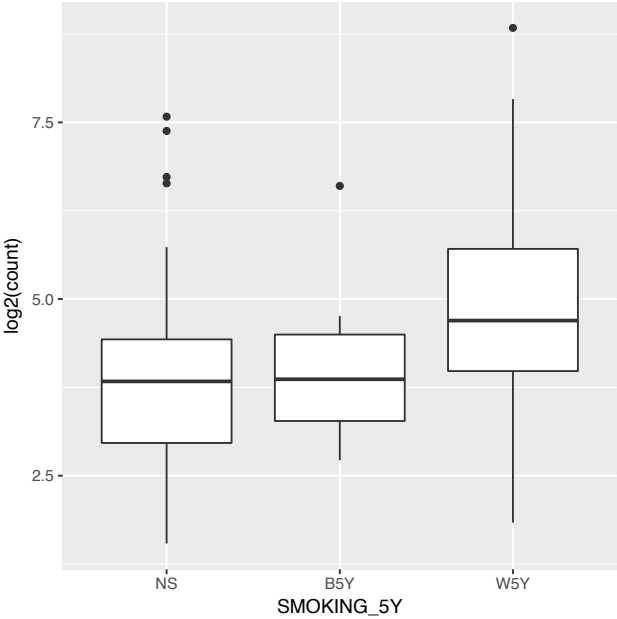
Supp. Fig. 8



Supp. Fig. 8 (continued)



Supp. Fig. 9



Supp. Fig. 10

