HGGA, Volume *5*

Supplemental information

Massively parallel reporter assay confirms regulatory potential of hQTLs and reveals important variants in lupus and other autoimmune diseases

Yao Fu, Jennifer A. Kelly, Jaanam Gopalakrishnan, Richard C. Pelikan, Kandice L. Tessneer, Satish Pasula, Kiely Grundahl, David A. Murphy, and Patrick M. Gaffney

Table of Contents

Figure S1. MPRA QC and descriptive statistics. Page 2

Figure S2. Tested variants located in ENCODE cREs. Page 3

Figure S3. emSeq/emVar variants on SLE risk haplotypes. Pages 4-7

Figure S4. Box plots of SLE emVar variants. Pages 8-11

Figure S5. Box plots of AI emVar variants. Page 12

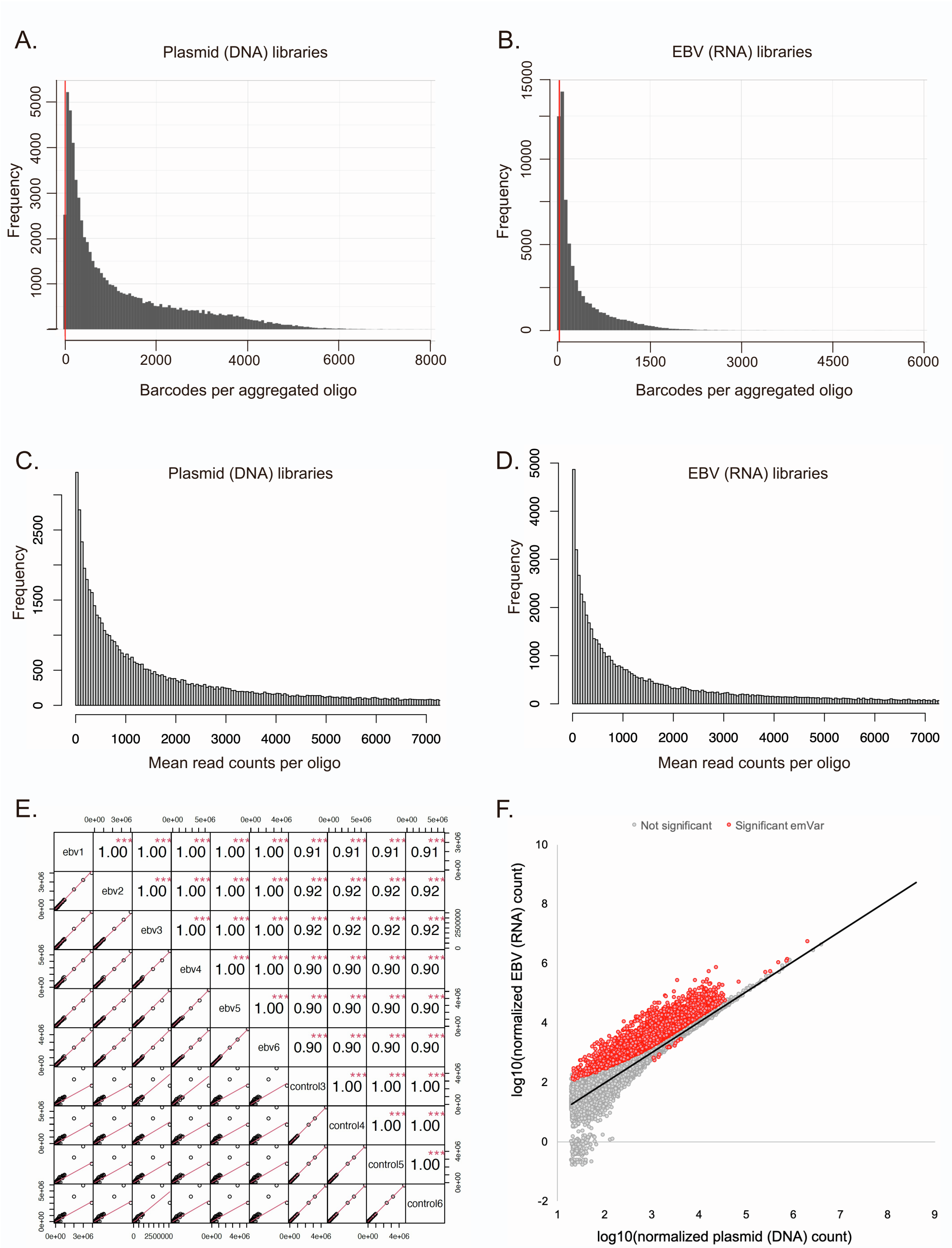


Figure S1. MPRA QC and descriptive statistics. A. Distribution of the number of oligos with > 10 barcodes (n = 66,131; 98.9%) in the aggregated plasmid control libraries. **B.** Distribution of the number of oligos with > 10 barcodes (n = 61,096, 91.3%) in the aggregated EBV replicate libraries. **C.** Distribution of the number of oligos with > 20 mean counts (n = 65,532; 98%) in the aggregated plasmid control libraries. **D.** Distribution of the number of oligos with > 20 mean counts (n = 64,361, 96.2%) in the aggregated EBV B replicate libraries. **E.** Correlation matrix of oligo counts for each replicate library. **F.** Scatterplot of pairwise comparisons of normalized oligo counts in aggregated plasmid replicates (x-axis) and aggregated EBV replicates (y-axis). Significant emVars are indicated in red.

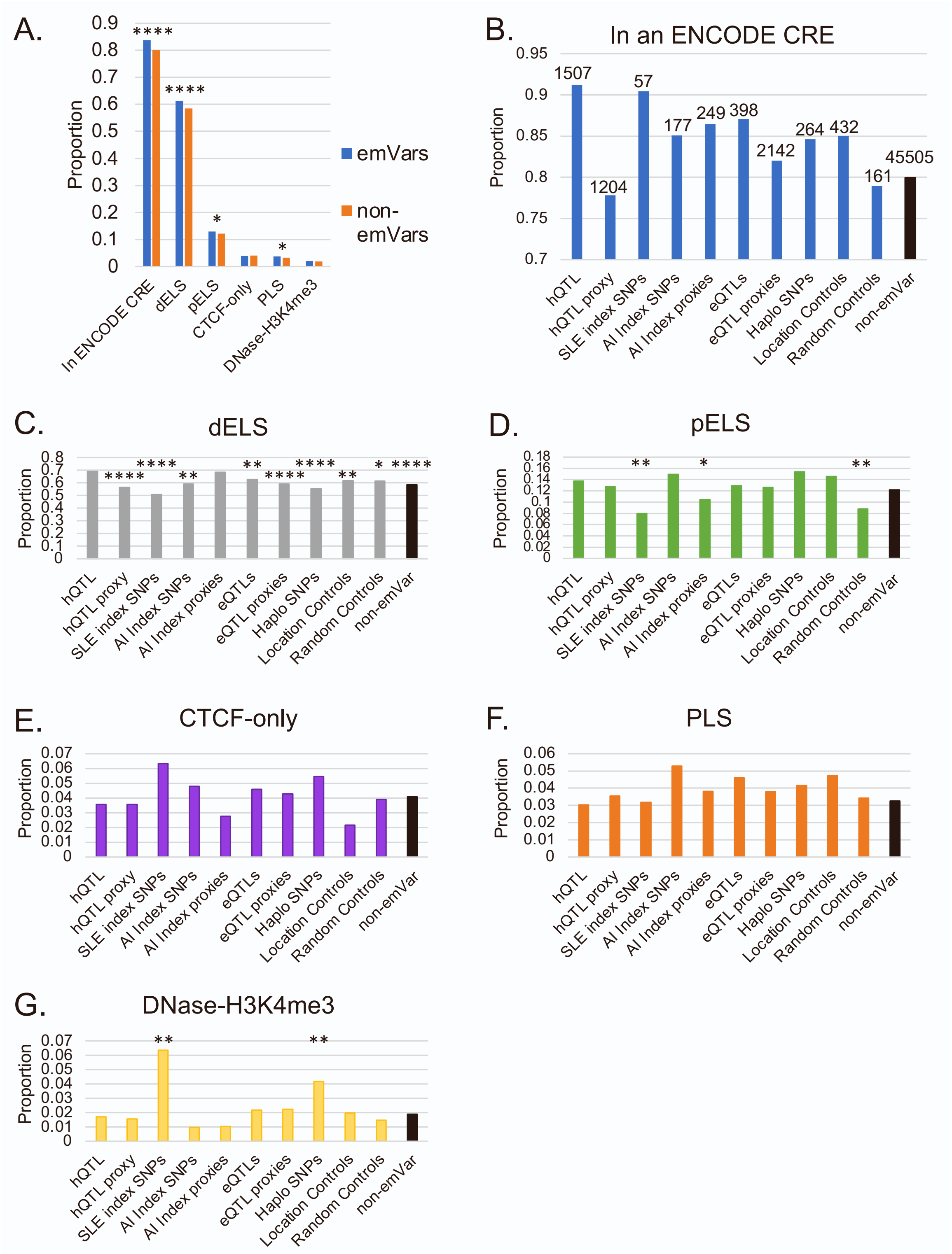
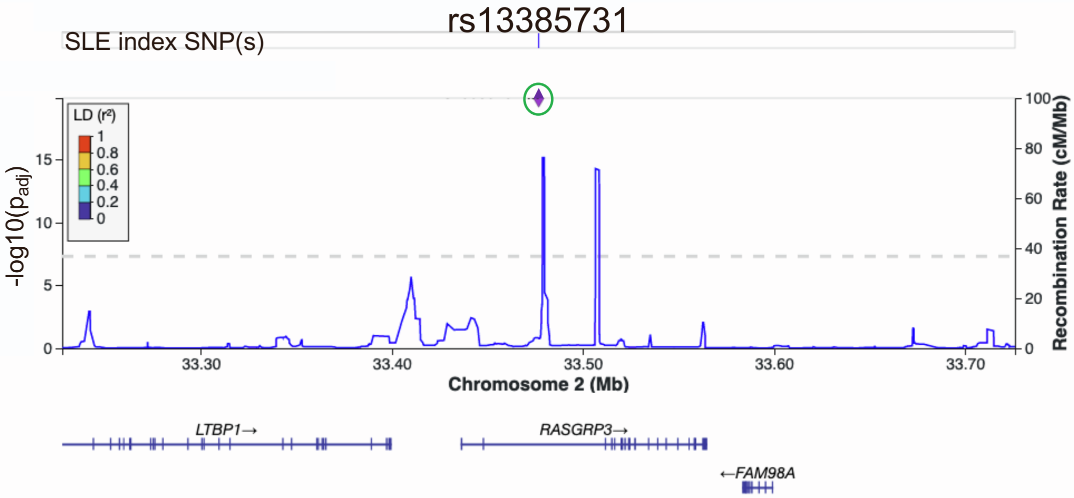
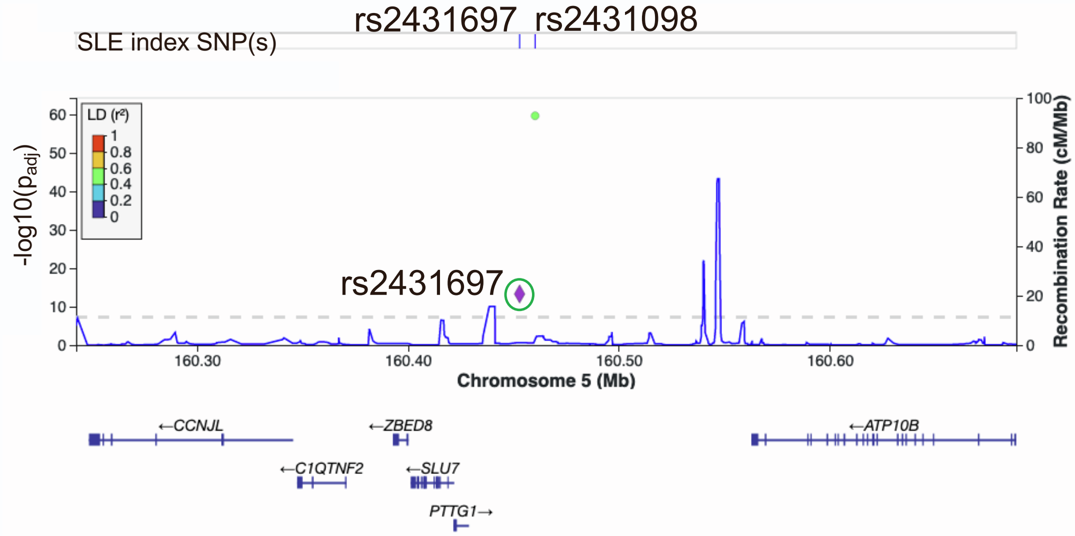


Figure S2. Tested variants located in ENCODE cREs. A. Proportion (y-axis) of emVars and non-emVars in the different types (x-axis) of cREs. **B.** Proportion (y-axis) of emVars (blue) in cREs by variant type (y-axis). Count is provided above each bar. **C.** Proportion of emVars (gray) in distal enhancer-like signatures by variant type. **D.** Proportion of emVars (green) in proximal enhancerlike signatures by variant type. **E.** Proportion of emVars (purple) in CTCF-only cREs by variant type. **F.** Proportion of emVars (orange) in promoter like signatures by variant type. **G.** Proportion of emVars (yellow) in DNase-H3K4me3 cREs by variant type. Significant differences in proportion of each cRE in hQTLs compared to each other variant type are shown: *<0.05; ** < 0.01, ***<0.001, ****<0.0001.

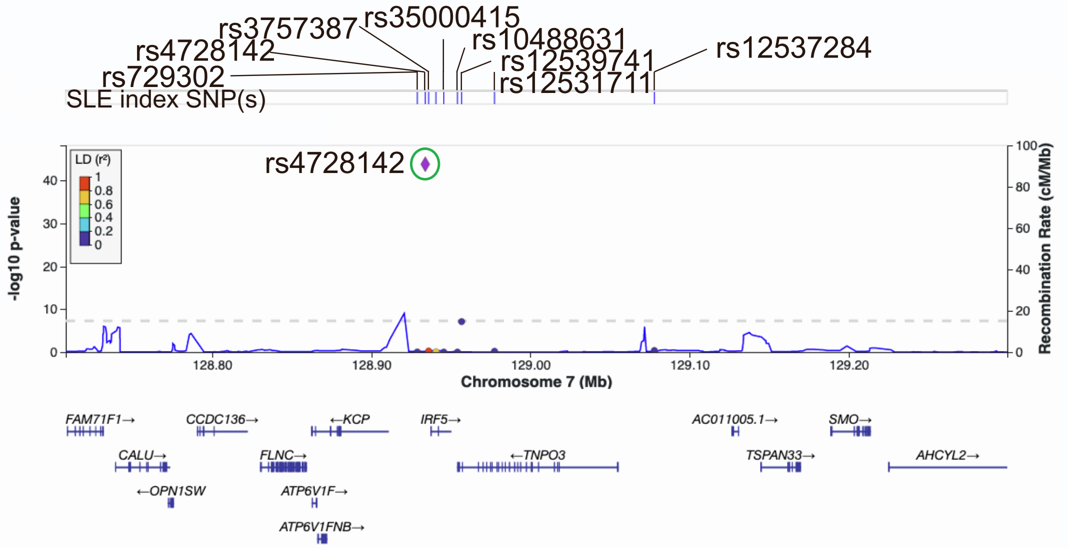
A. RASGRP3 – chr2:33227836-33726592



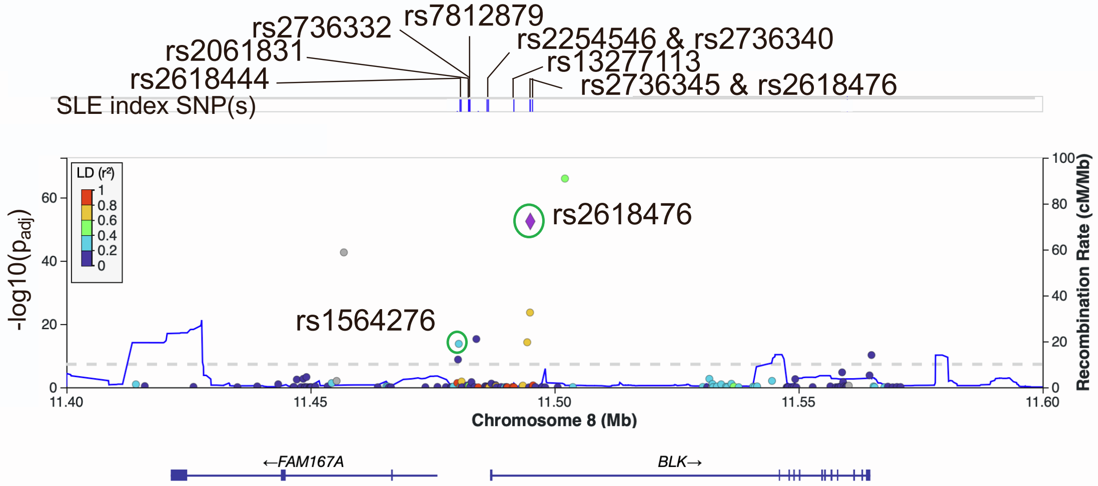
B. PTTG1 – miRNA146 – chr5:160243289-160688961



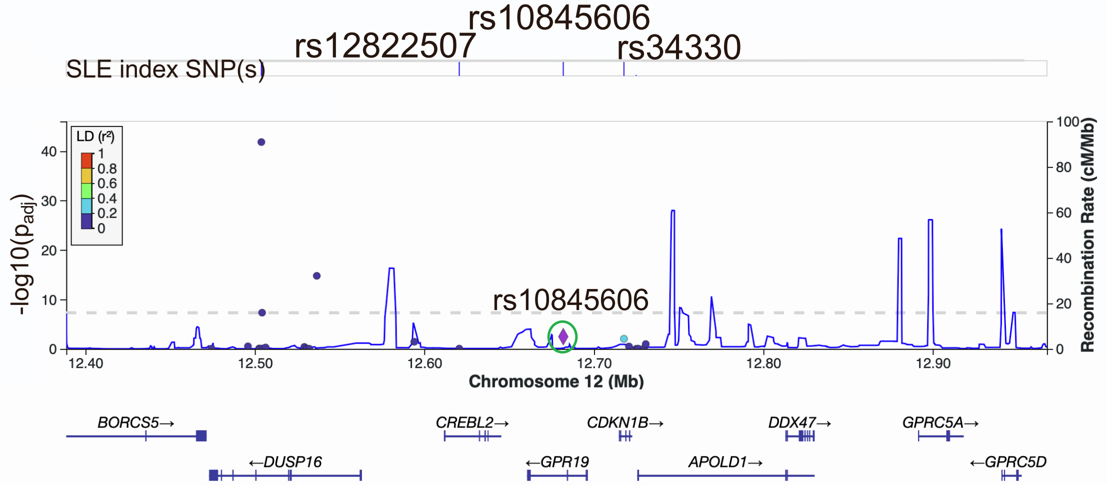
C. IRF5 – chr7:128708061-129299605



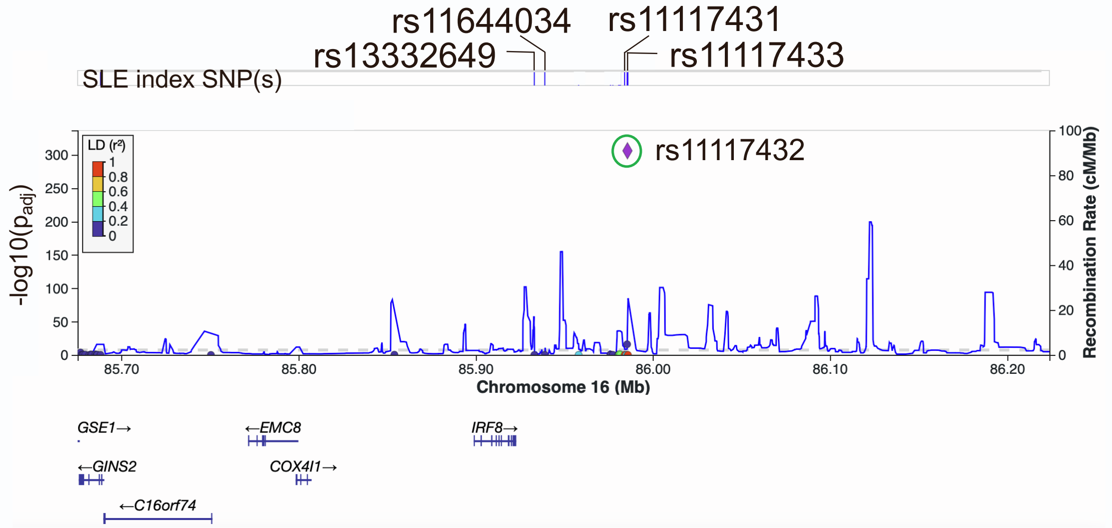
D. BLK – chr8:11400000-11600000



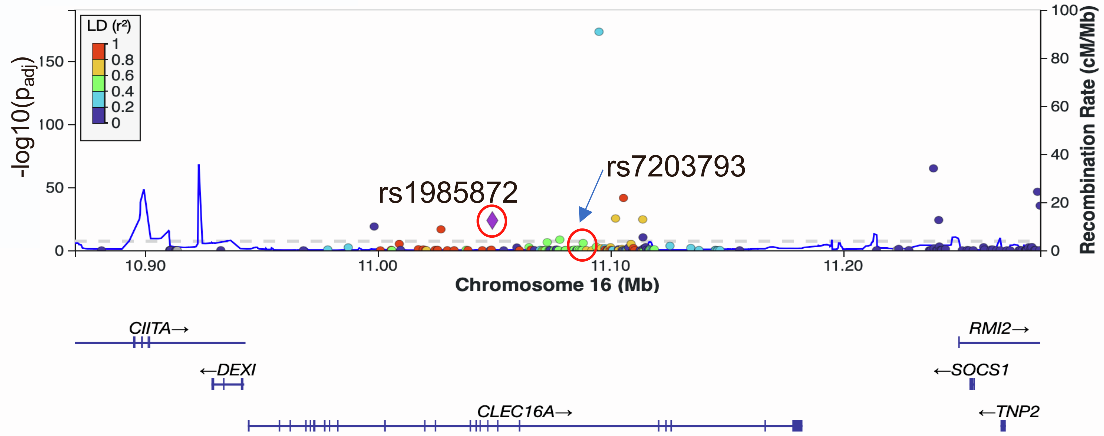
E. CREBL2 - GPR19 - CDKN1B – chr12:12388701-12967318



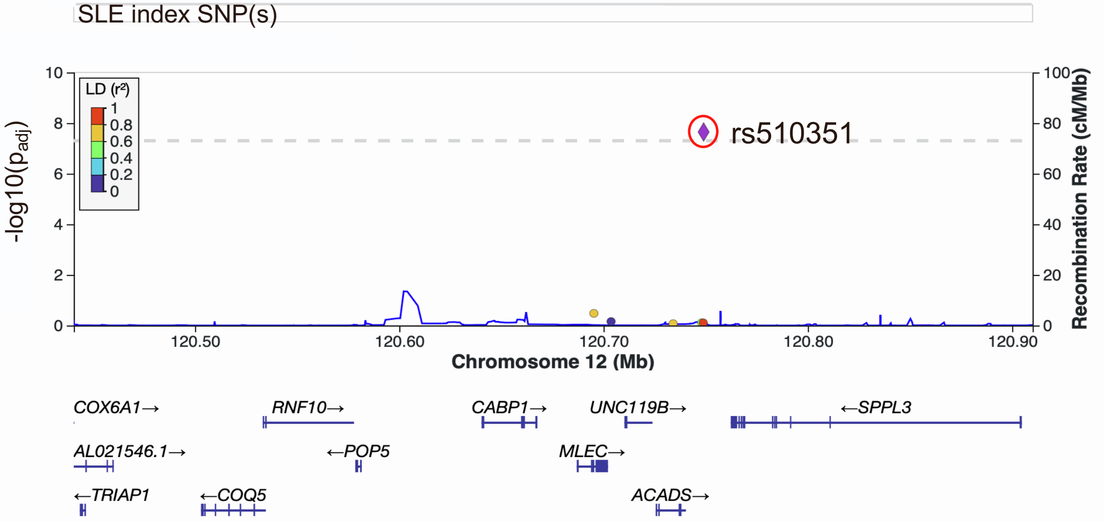
F. IRF8 – chr16:85675256-86223875



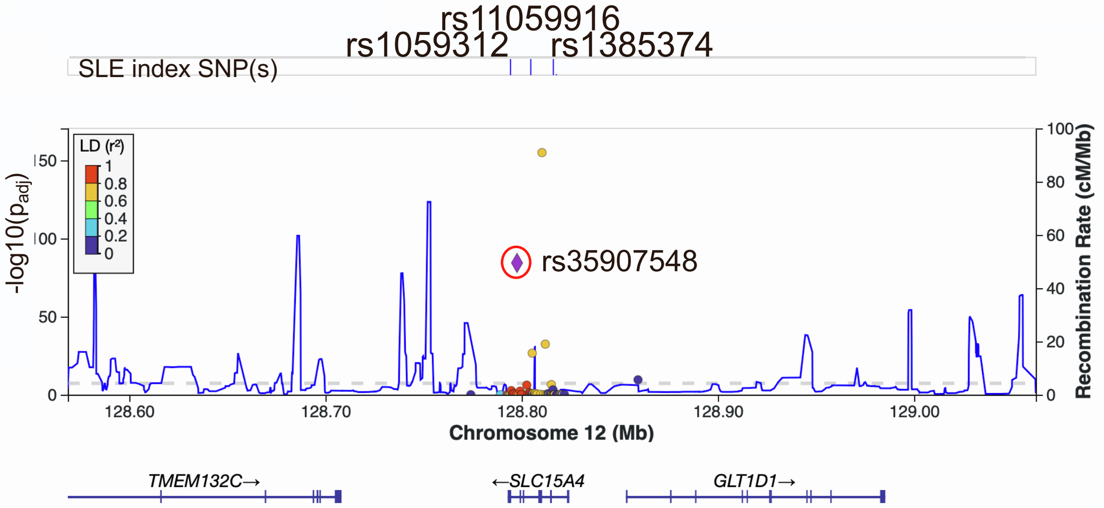
G. CIITA – CLEC16A – SOCS1 – chr16: 10870000-11284542

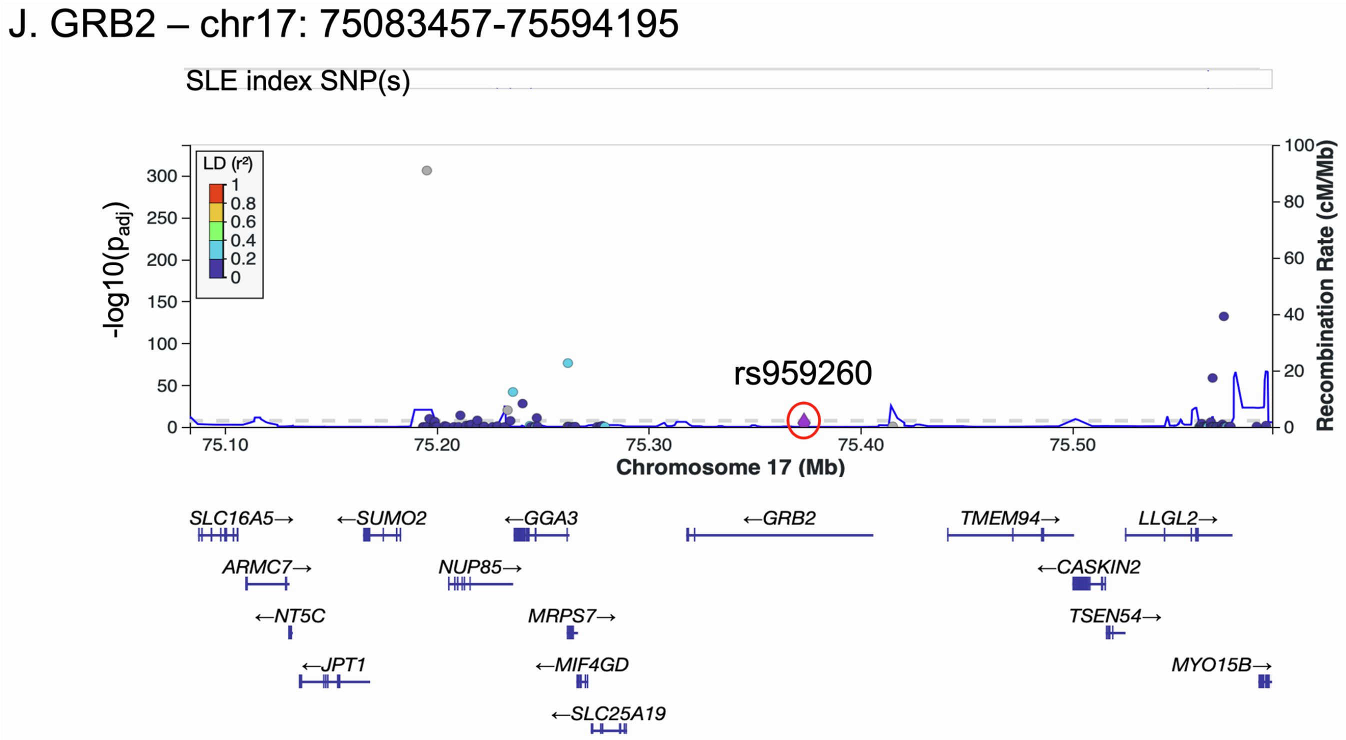


H. CABP1 – SPPL3 – chr12: 120440727-120909910



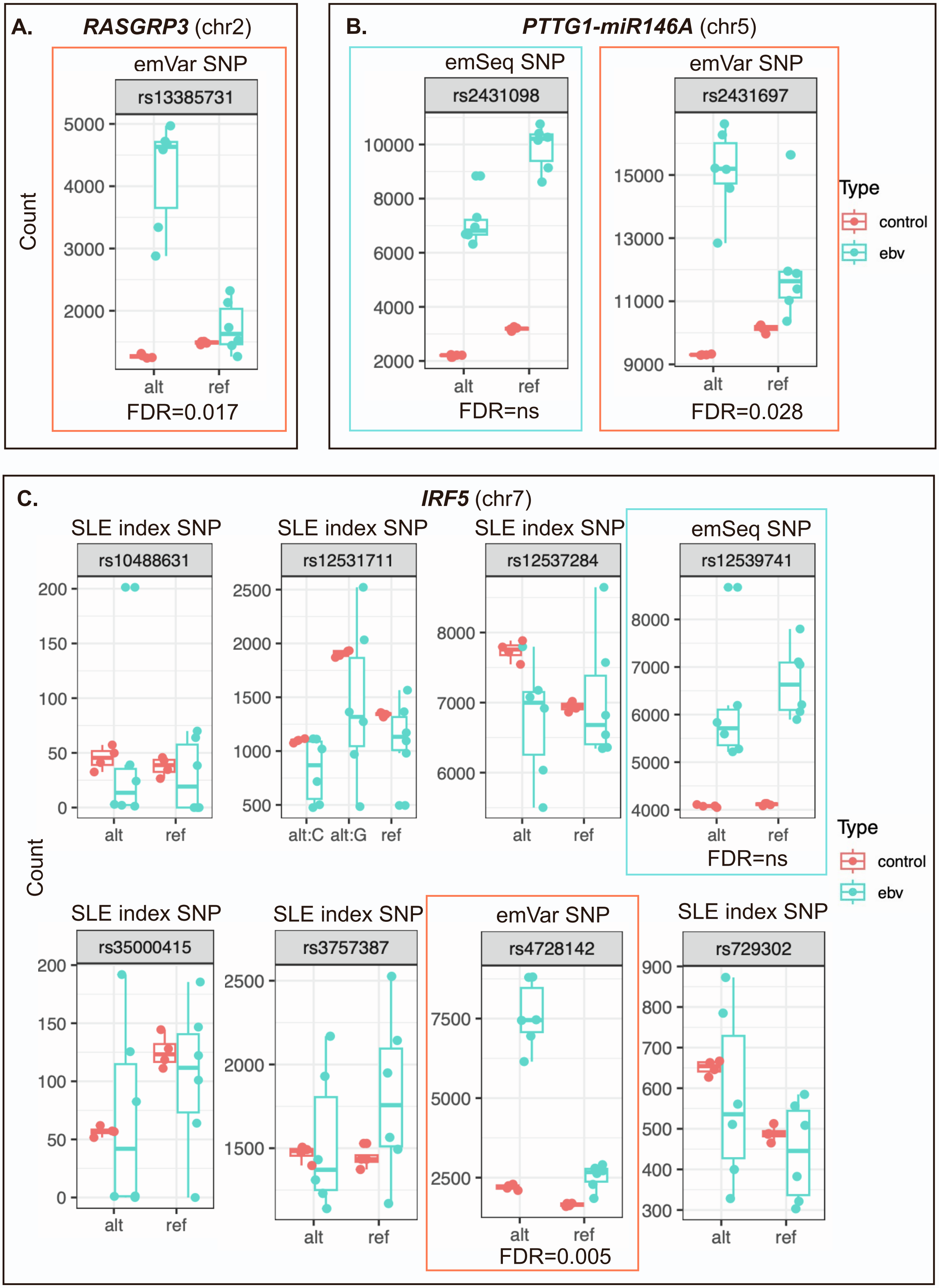
I. SLC15A4 – chr12: 128568772-1290619

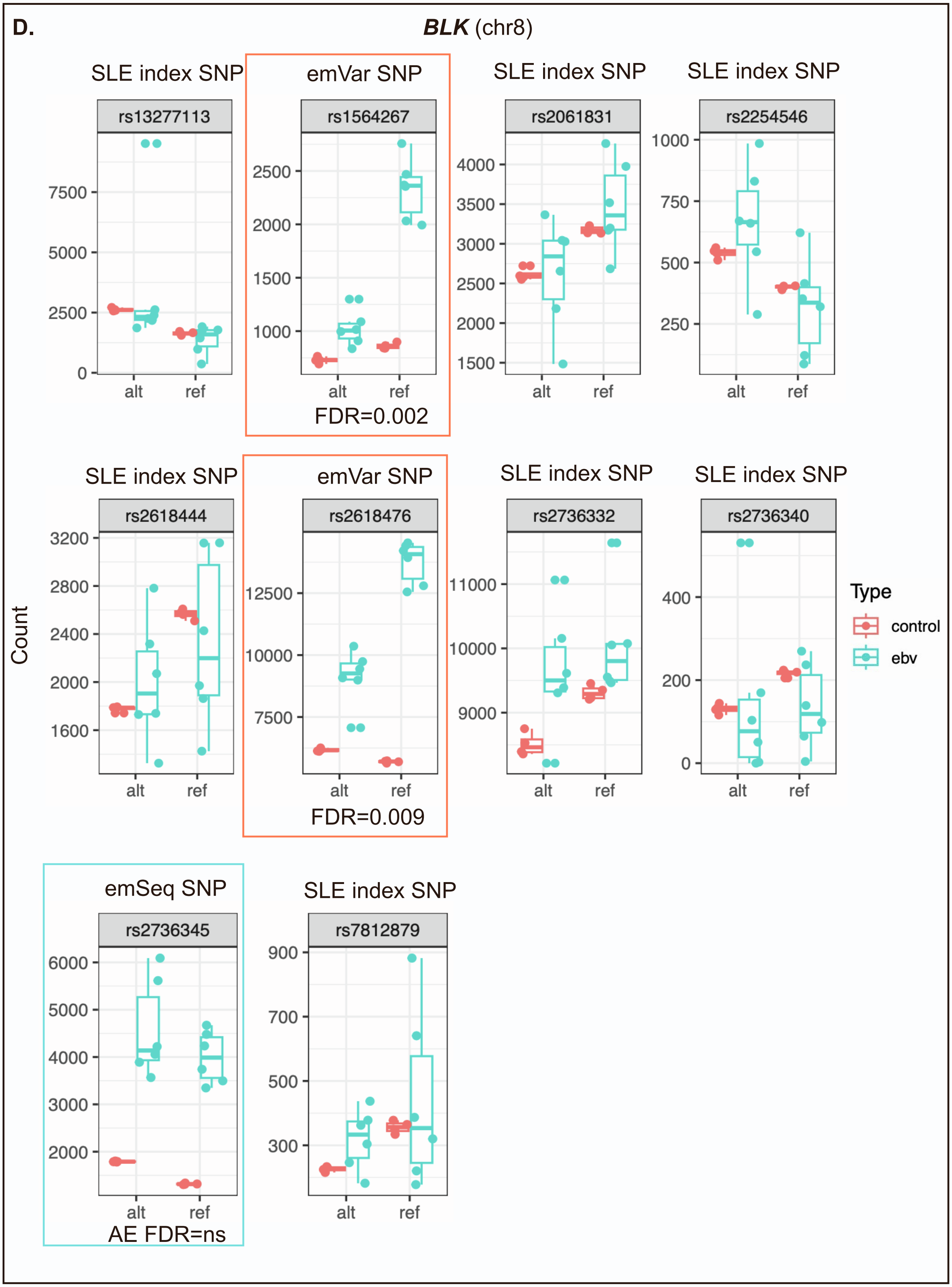


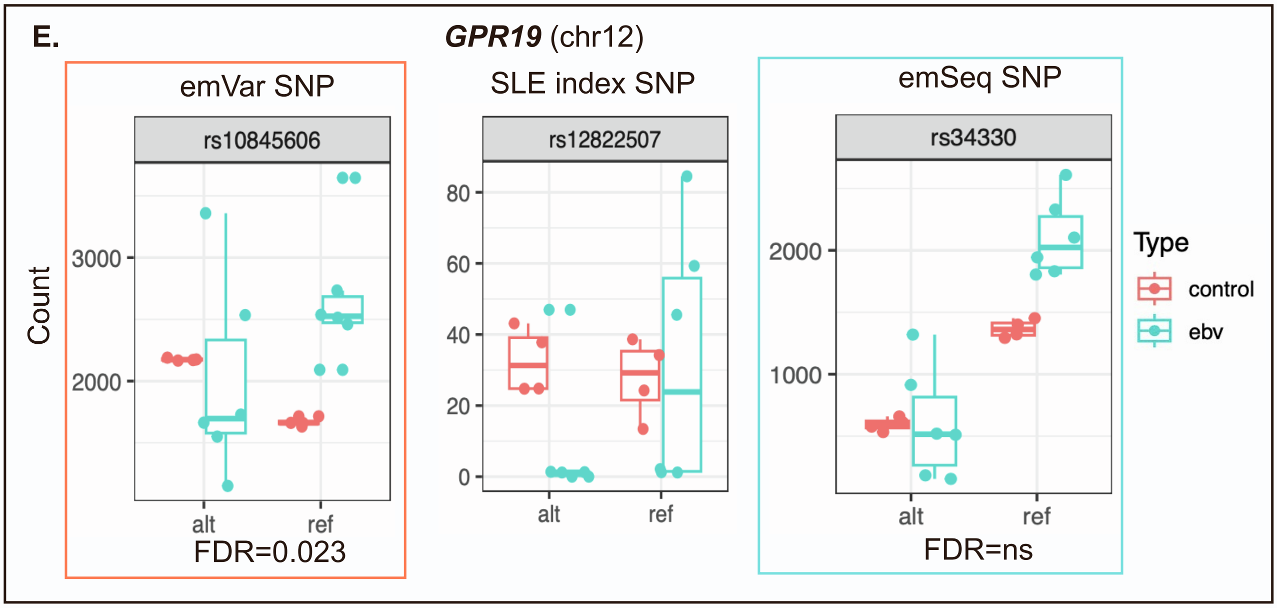


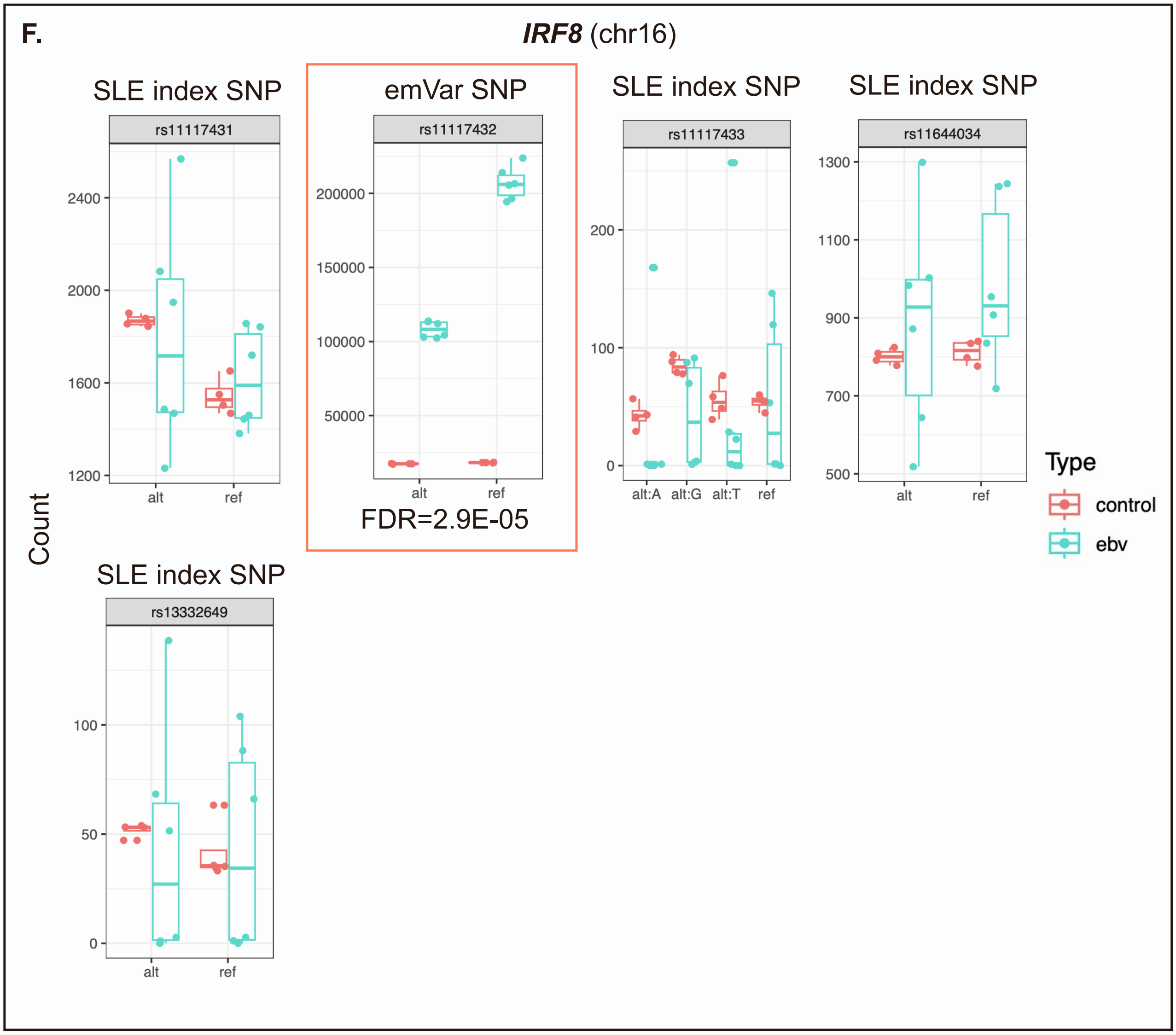
J. GRB2 – chr17: 75083457-75594195

Figure S3. emVar/AE variants on SLE risk haplotypes. LocusZoom plots demonstrating emSeq and emVar effects on SLE risk haplotypes. Evaluated index SNPs are presented at the top of graph. Variants evaluated, their genomic location, and genes in the region are plotted on the x-axis. emVars are represented as a purple diamond. Published SLE index SNP emVars are circled in green **(A-F)** and novel SLE emVars are circled in red **(G-J)**. Variants are colored based on their LD r^2 values with the circled emVar (see LD key). The -$\log_{10}(p_{adj})$ of the emSeq score for each variant is plotted on the y-axis.


A.
RASGRP3 (chr2)
emVar SNP
rs13385731
Count
5000
4000
3000
2000
alt
ref
FDR=0.017
B.
PTTG1-miR146A (chr5)
emSeq SNP
rs2431098
10000
8000
6000
4000
2000
alt
ref
FDR=ns
emVar SNP
rs2431697
15000
13000
11000
9000
alt
ref
FDR=0.028
Type
control
ebv
C.
IRF5 (chr7)
SLE index SNP
rs10488631
200
150
100
50
0
alt
ref
SLE index SNP
rs12531711
2500
2000
1500
1000
500
alt:C
alt:G
ref
SLE index SNP
rs12537284
8000
7000
6000
alt
ref
emSeq SNP
rs12539741
8000
7000
6000
5000
4000
alt
ref
FDR=ns
Type
control
ebv
Count
SLE index SNP
rs35000415
200
150
100
50
0
alt
ref
SLE index SNP
rs3757387
2500
2000
1500
alt
ref
emVar SNP
rs4728142
7500
5000
2500
alt
ref
FDR=0.005
SLE index SNP
rs729302
900
800
700
600
500
400
300
alt
ref


D.
BLK (chr8)
SLE index SNP
rs13277113
emVar SNP
rs1564267
FDR=0.002
SLE index SNP
rs2061831
SLE index SNP
rs2254546
SLE index SNP
rs2618444
emVar SNP
rs2618476
FDR=0.009
SLE index SNP
rs2736332
SLE index SNP
rs2736340
emSeq SNP
rs2736345
AE FDR=ns
SLE index SNP
rs7812879
Count
alt
ref
Type
control
ebv
0
2500
5000
7500
1000
1500
2000
2500
1500
2000
2500
3000
3500
4000
250
500
750
1000
1600
2000
2400
2800
3200
7500
10000
12500
9000
10000
11000
0
200
400
2000
3000
4000
5000
6000
300
500
700
900


E.
GPR19 (chr12)
emVar SNP
rs10845606
SLE index SNP
rs12822507
emSeq SNP
rs34330
Count
3000
2000
80
60
40
20
0
2000
1000
alt
ref
FDR=0.023
FDR=ns
Type
control
ebv


F.
IRF8 (chr16)
SLE index SNP
rs11117431
emVar SNP
rs11117432
SLE index SNP
rs11117433
SLE index SNP
rs11644034
SLE index SNP
rs13332649
Count
2400
2000
1600
1200
200000
150000
100000
50000
200
100
0
1300
1100
900
700
500
100
50
0
alt
ref
alt:A
alt:G
alt:T
FDR=2.9E-05
Type
control
ebv

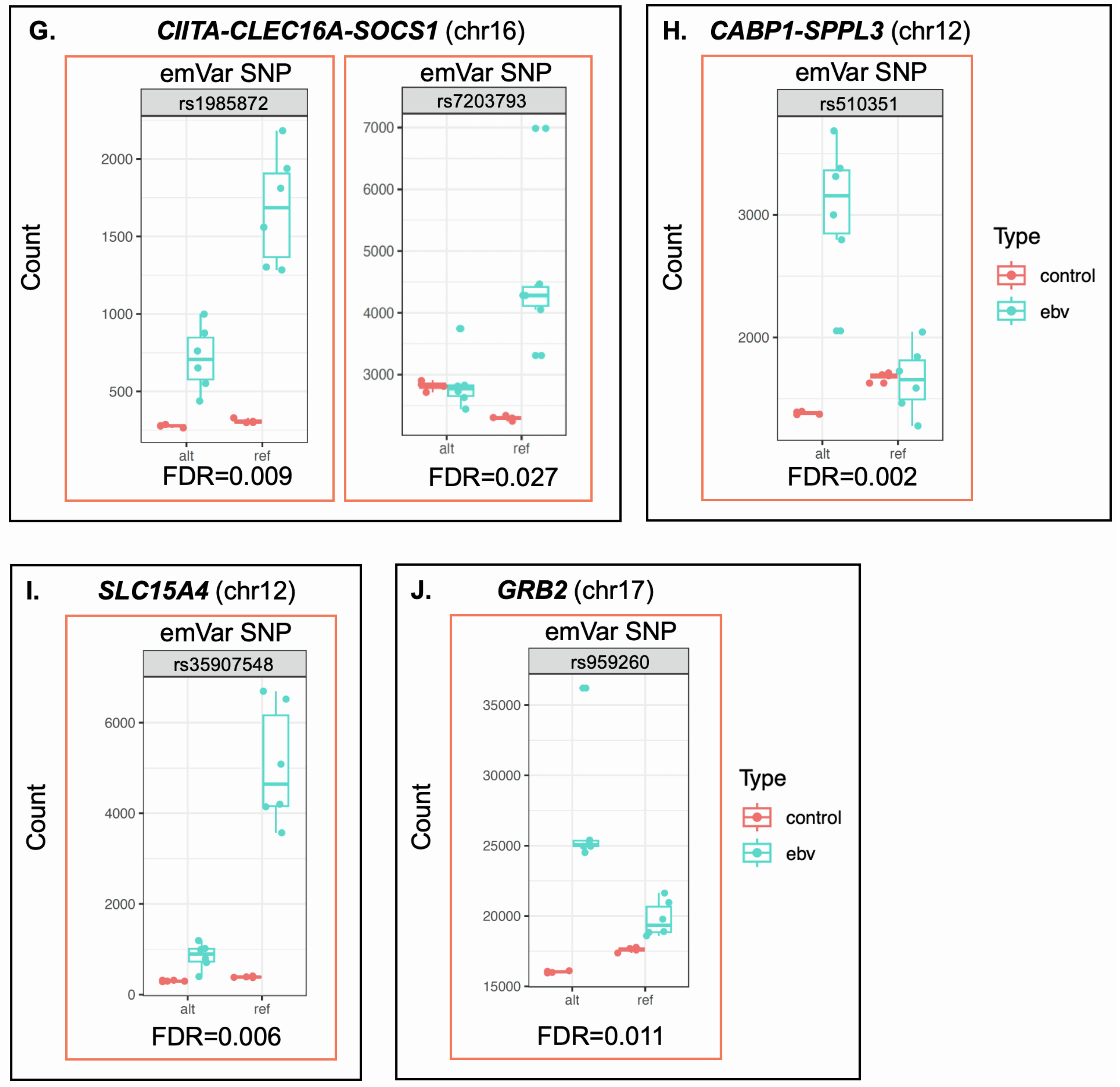


Figure S4. Box plots of SLE allelic effect variants. Box plots of normalized counts for EBV B replicates (green) and controls (orange) at each AI allelic variant. Count is plotted on the y-axis and allele (ref/alt) is plotted on the x-axis. The FDR q value and risk gene are provided. EmSeqs are boxed in green and emVars are boxed in orange.

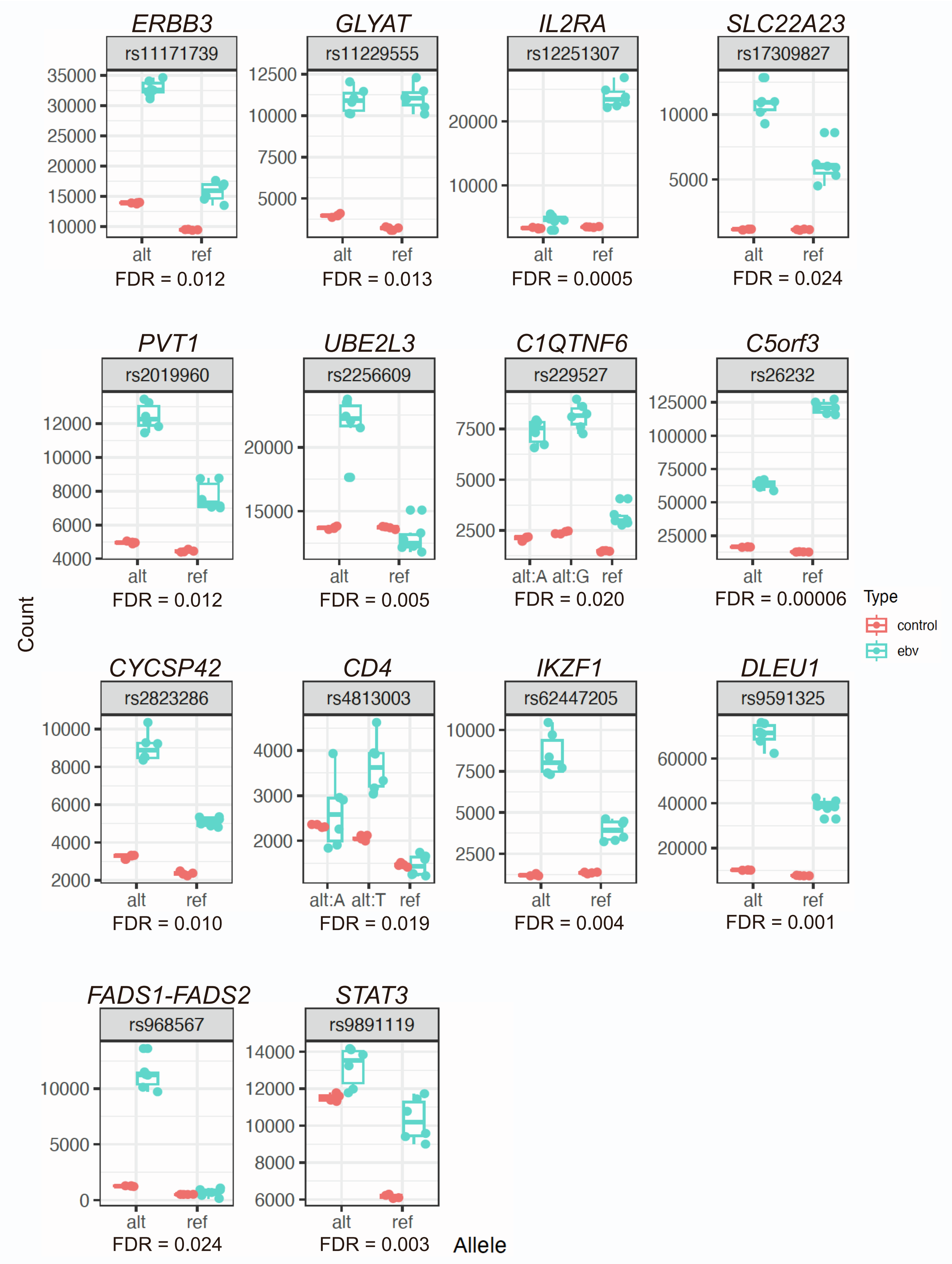


Figure S5. Box plots of AI emVars. Box plots of normalized counts for EBV B replicates (green) and controls (orange) at each emVar. Count is plotted on the y-axis and allele (ref/alt) is plotted on the x-axis. The FDR q value and risk gene are provided. Panels are ordered by rsID.

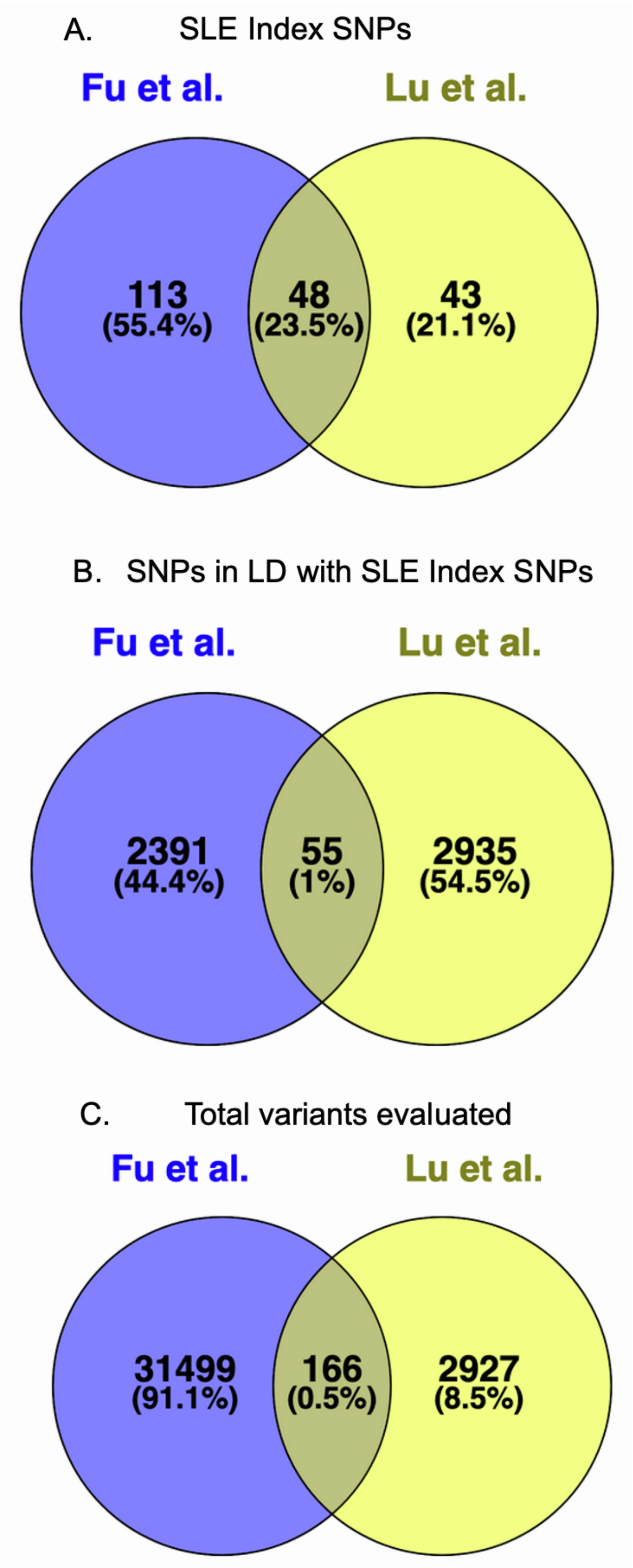


Figure S6. Venn diagrams displaying overlap between our study (Fu et al.) and an SLE MPRA study by Lu et al [13]. (A) Overlap of SLE index SNPs evaluated by both studies. (**B)** Overlap of SNPs in LD with SLE index SNPs evaluated by both studies. (**C)** Overlap of total variants evaluated by both studies.