Current Biology, Volume *31*

Supplemental Information

An ancient viral epidemic involving host coronavirus interacting genes more than 20,000 years ago in East Asia

Yassine Souilmi, M. Elise Lauterbur, Ray Tobler, Christian D. Huber, Angad S. Johar, Shayli Varasteh Moradi, Wayne A. Johnston, Nevan J. Krogan, Kirill Alexandrov, and David Enard

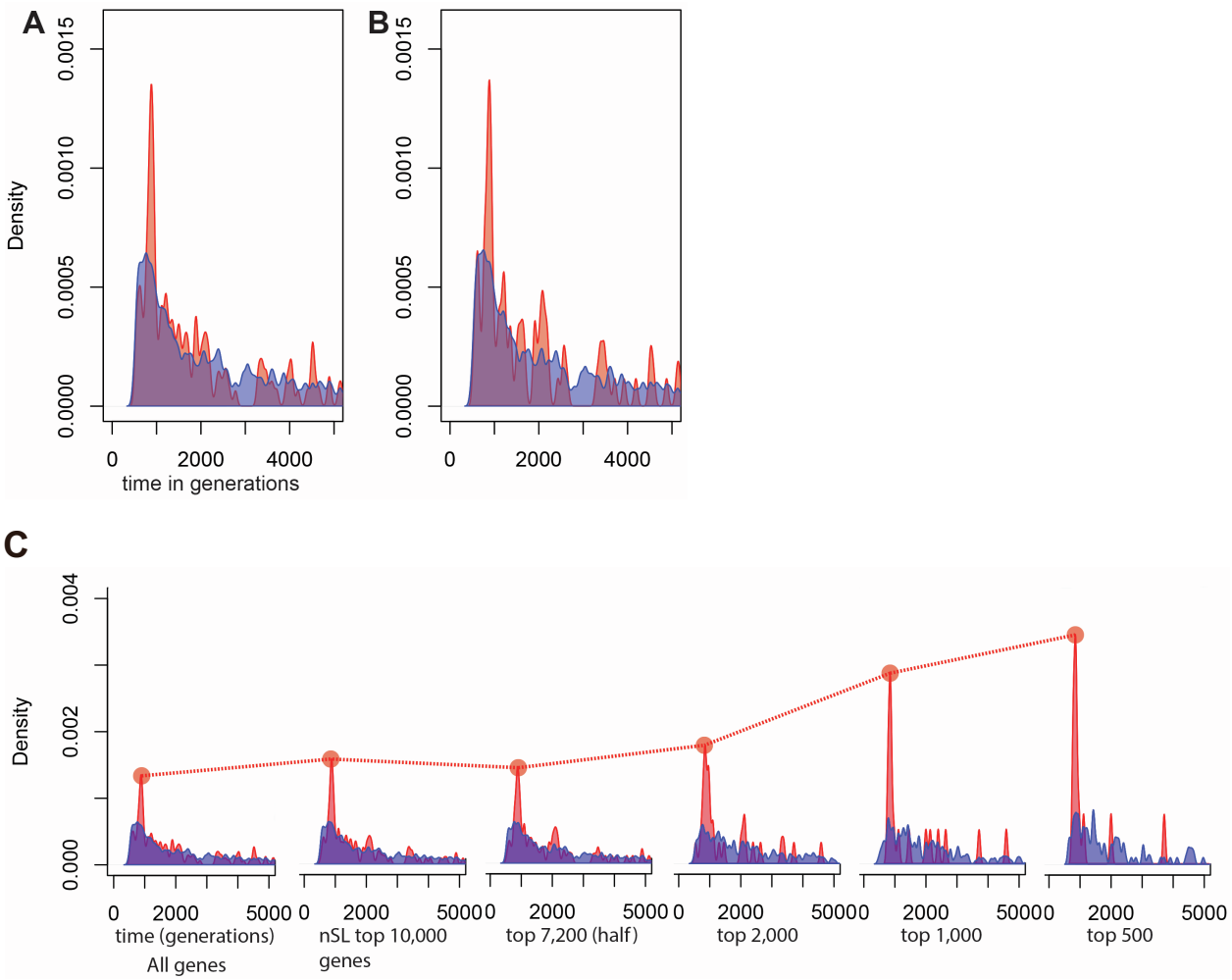


Figure S1. Timing of selection start at CoV-VIPs, with or without removing GO functions with a significant peak between 770 and 970 generations. Related to STAR Methods and Figure 2.

Same legend as Figure 2. A) All CoV-VIPs. B) CoV-VIPs with at least one of the 16 GO functions with a significant peak between 770 and 970 generations ago are excluded (31% of VIPs; Data S1F). Related to Figure 2. C) The figure shows the amplitude of the peak of selection start times for increasingly high nSL thresholds. For example, for the nSL top 1,000, only selection start times at genes within the top 1,000 nSL (average rank over East Asian populations, lower rank of the 1Mb and 2Mb nSL windows) are included to get the pink and blue distributions. Related to Figure 2.

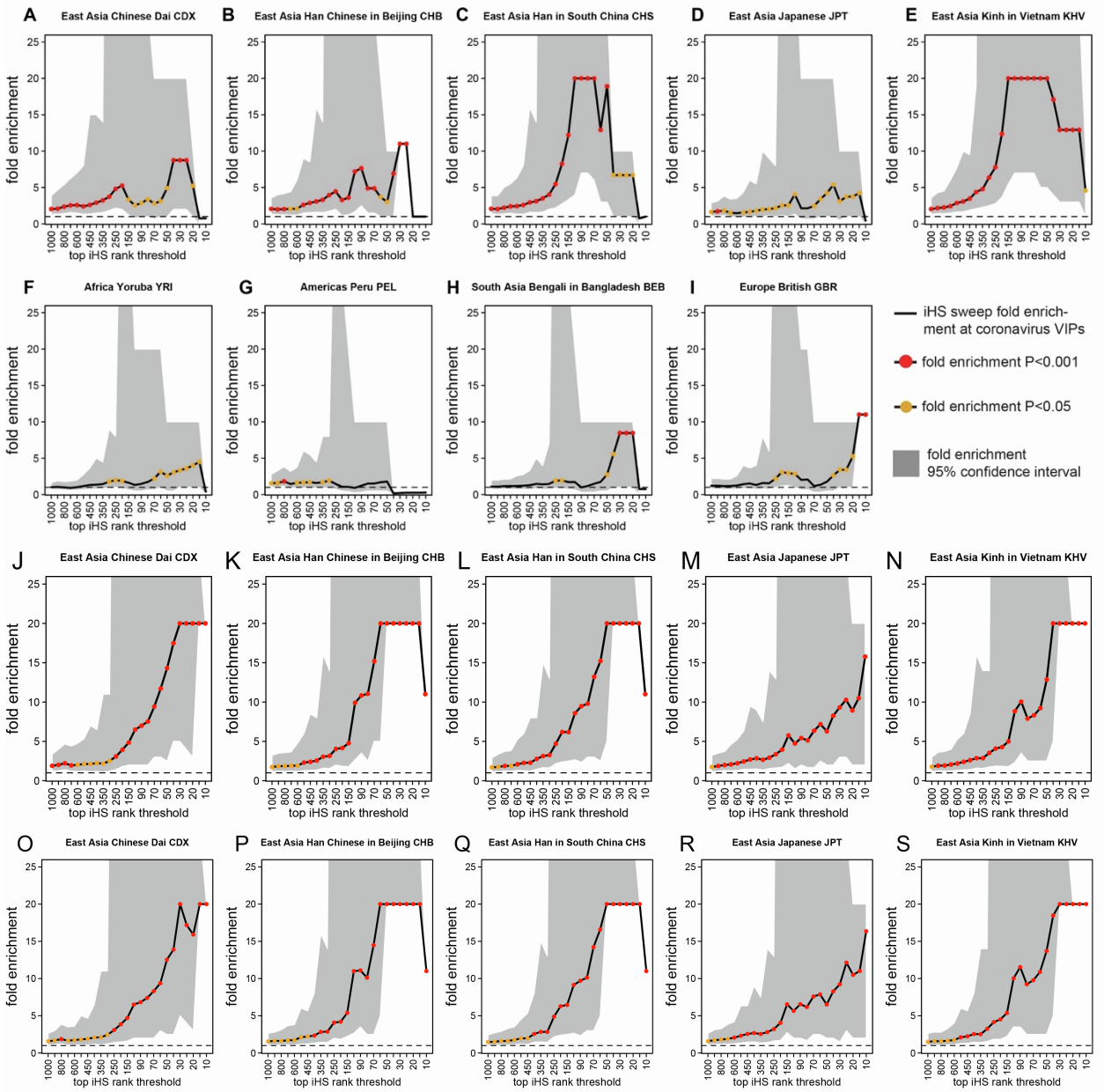


Figure S2. CoV-VIPs sweep enrichment with his. Related to STAR Methods and Figure 1. Related to Figure 1. Same legend as Figure 1. A) to I) The only change compared to Figure 1 is the use of iHS instead of nSL. J) to N) Same as Figure 1(nSL), but no matching for confounding factors. O) to S) Same as Figure 1 (nSL), but also matching for SNP density in addition to all the other confounding factors.

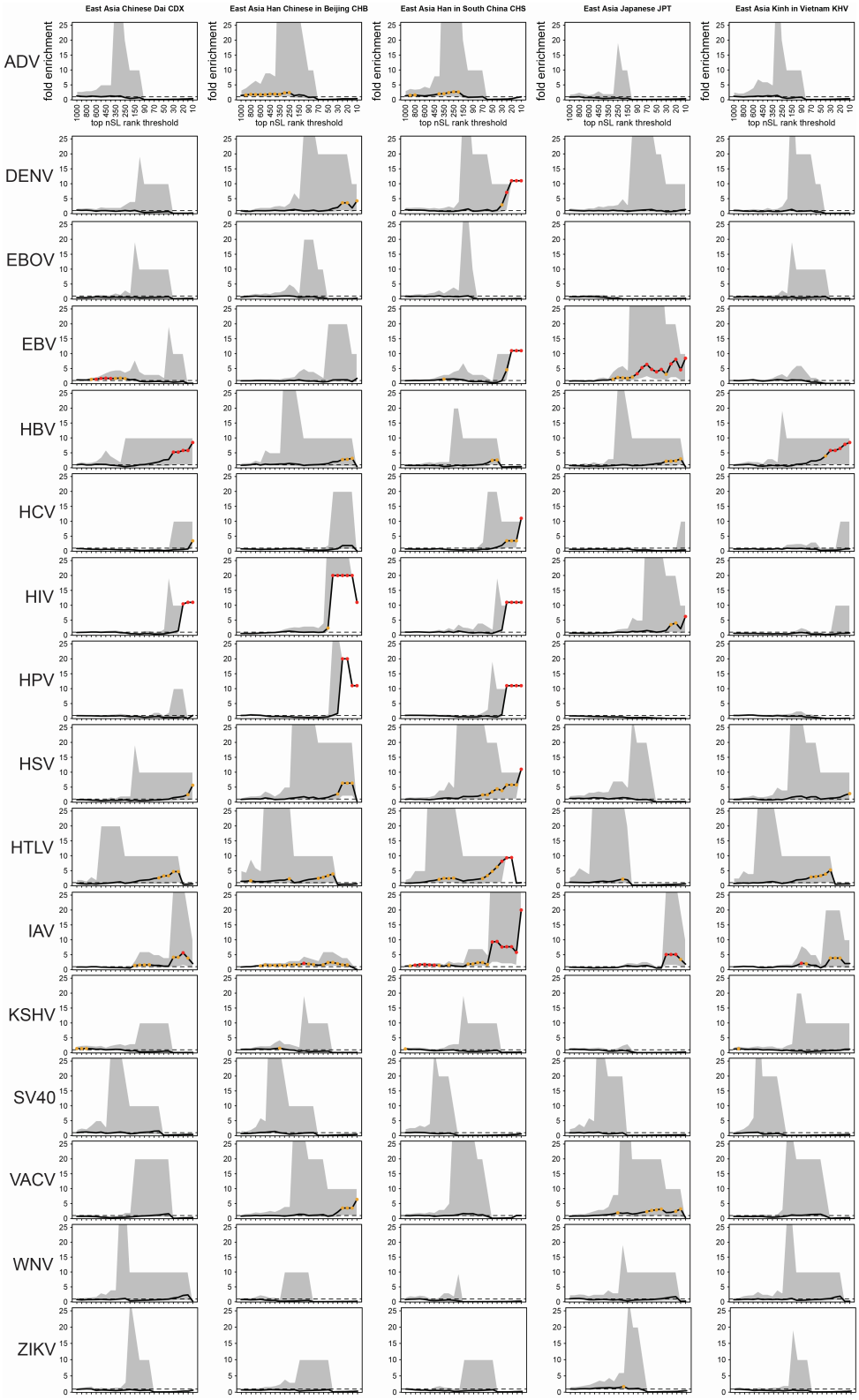


Figure S3. nSL sweep enrichment curves for 17 other viruses in East Asia. Related to STAR Methods and Figure 1.

Same legend as in Figure 1. Whole curve P>0.05 for all viruses. Related to Figure 1.

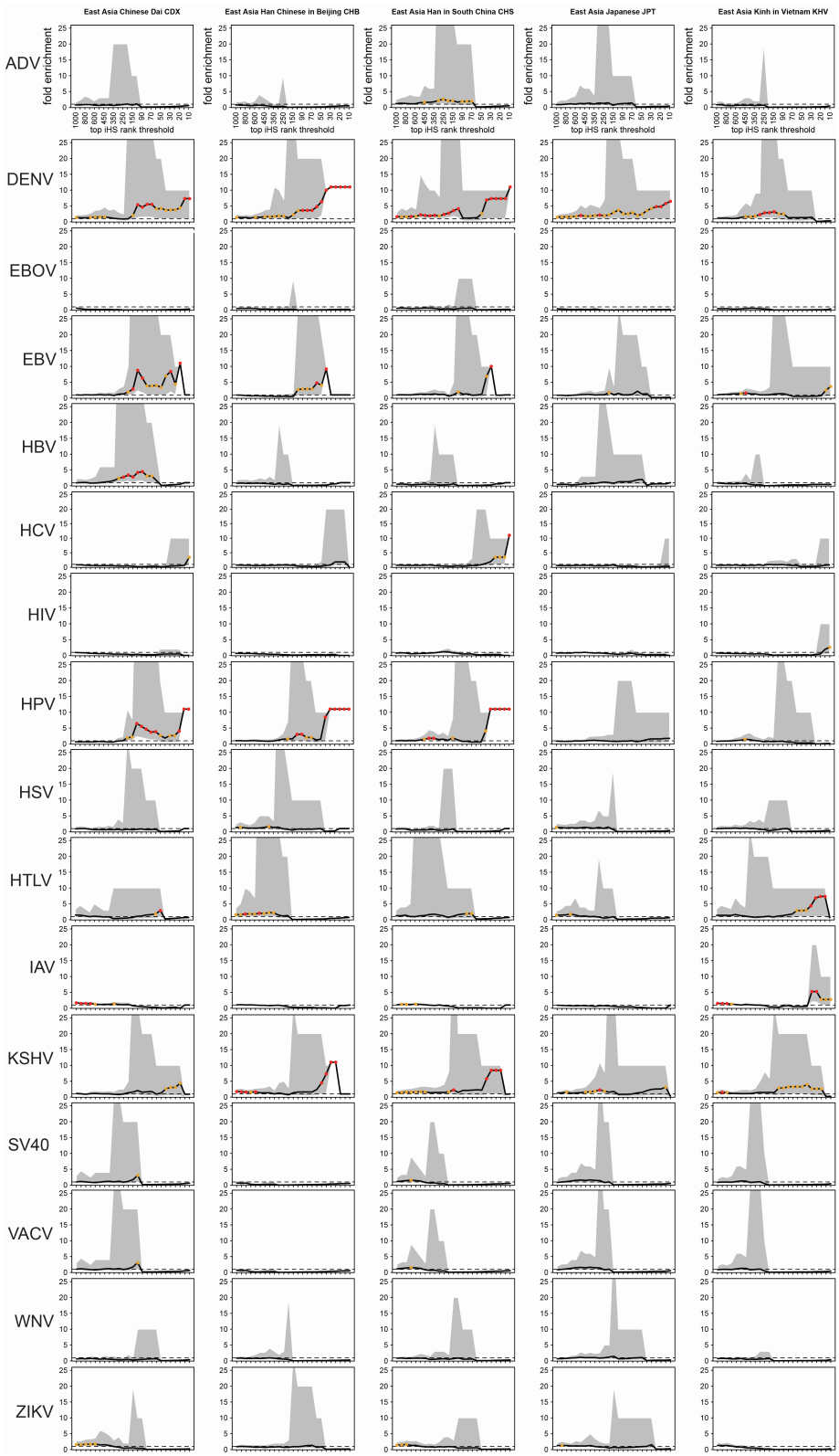
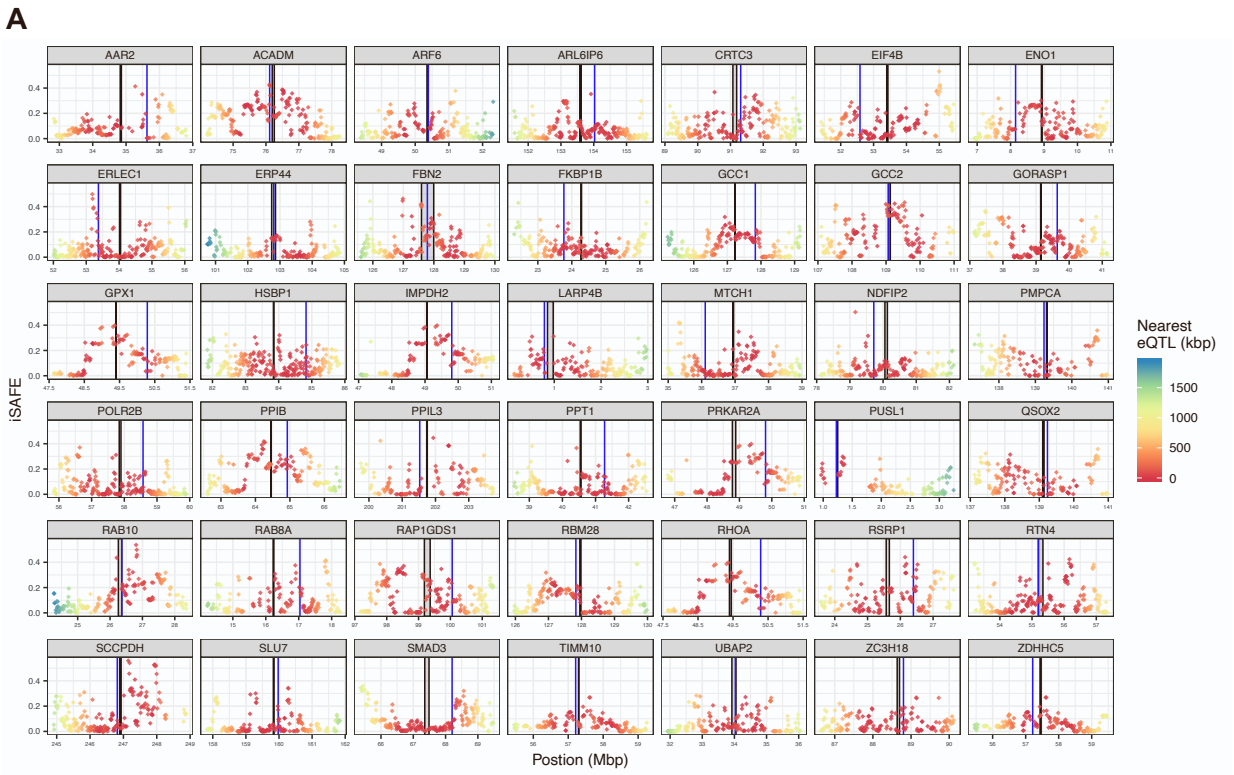


Figure S4. iHS sweep enrichment curves for 17 other viruses in East Asia. Related to STAR Methods and Figure 1.

Same legend as in Figure 1. Whole curve P>0.05 for all viruses. Related to Figure 1.


A
AAR2
ACADM
ARF6
ARL6IP6
CRTC3
EIF4B
ENO1
ERLEC1
ERP44
FBN2
FKBP1B
GCC1
GCC2
GORASP1
GPX1
HSBP1
IMPDH2
LARP4B
MTCH1
NDFIP2
PMPCA
POLR2B
PPIB
PPIL3
PPT1
PRKAR2A
PUSL1
QSOX2
RAB10
RAB8A
RAP1GDS1
RBM28
RHOA
RSRP1
RTN4
SCCPDH
SLU7
SMAD3
TIMM10
UBAP2
ZC3H18
ZDHHC5
iSAFE
Postion (Mbp)
Nearest eQTL (kbp)
1500
1000
500
0

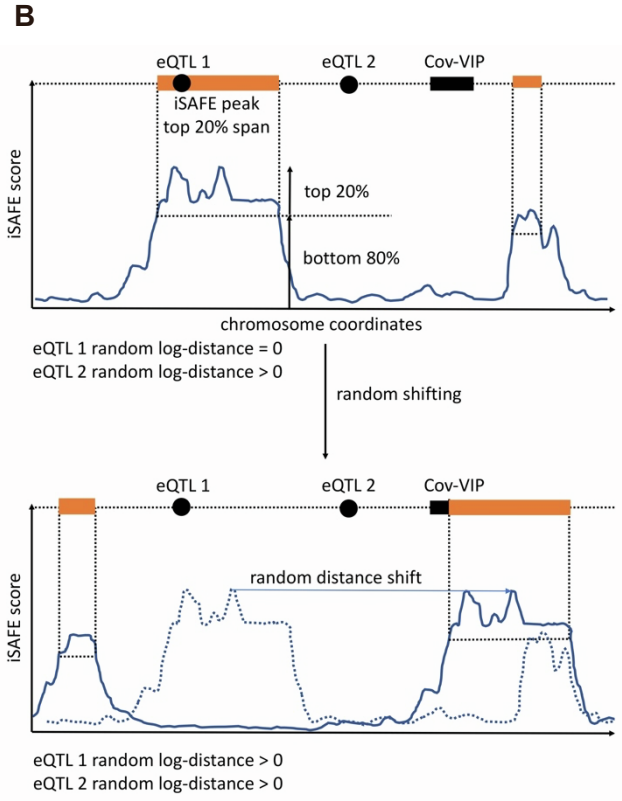
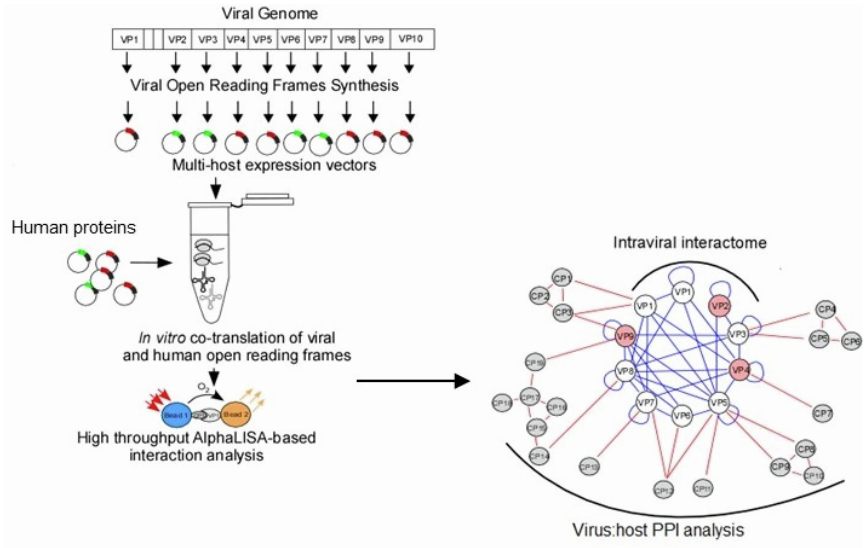

B
eQTL 1
eQTL 2
Cov-VIP
iSAFE peak
top 20% span
top 20%
bottom 80%
iSAFE score
chromosome coordinates
eQTL 1 random log-distance = 0
eQTL 2 random log-distance > 0
random shifting
eQTL 1
eQTL 2
Cov-VIP
random distance shift
iSAFE score
eQTL 1 random log-distance > 0
eQTL 2 random log-distance > 0

Figure S5. iSAFE peaks, GTEx eQTLs, Relate selected variants locations, and proximity test schematic. Related to STAR Methods and Figure 7.
Related to Figure 7. A) Dark lines: gene starts and gene ends. Blue line: Relate selected variant location. The color scale provides information about distance to the nearest GTEx eQTL. iSAFE peaks are not always clean, sharp peaks, and the Relate selected variants do not always overlap local iSAFE peaks, possibly as a result of both recombination since strong selection stopped, and weaker selection in more recent times (Figure 4). This is suggested by multiple steep iSAFE drops in the middle of peaks, as visible for example for ARL6IP6 at coordinate 153Mb. B) Sliding of iSAFE coordinates for the proximity ratio test
Black rectangle: area between the transcription start and end of a CoV-VIP. Black dot: coordinate of eQTL for the corresponding CoV-VIP. Orange area: area where distance between the iSAFE peak area and the closest eQTL is counted as zero. If the eQTL falls outside of an orange area, the distance is counted as distance to closest orange area edge. Dashed blue line in thre lower panel: original location of the real iSAFE score before random sliding.

A



B

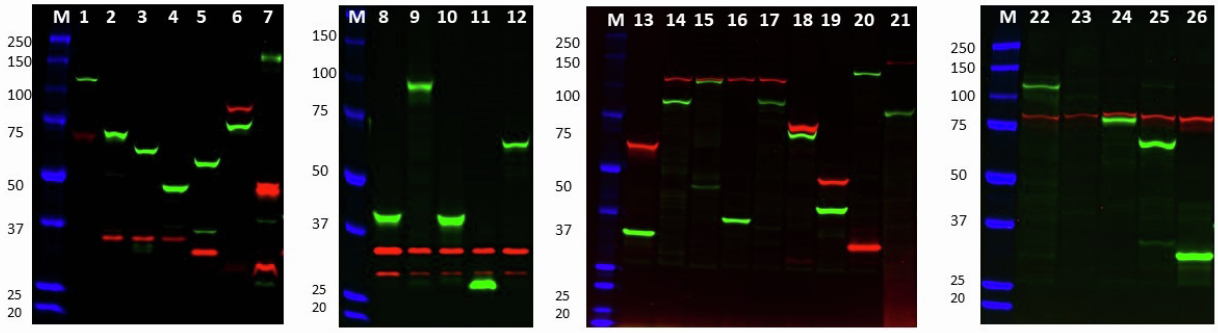


Figure S6. Schematic figure of in vitro expressed protein:protein interaction platform, and SDS-PAGE analysis of the LTE expressed SARS-COV-2 and 26 human proteins (the seven other tested CoV-VIPs are in Figure 6A). Related to STAR Methods and Figure 6.
A) Following co-expression of protein pairs in LTE system, the reactions are incubated with AlphaLISA beads. The interacting proteins are captured with streptavidin coated donor beads coupled to anti-mCherry nanobody and anti-GFP antibody acceptor beads. Upon protein: protein interaction, the acceptor bead comes to the proximity of donor bead. The singlet oxygen produced by donor beads reacts with thioxine derivative in the acceptor bead and subsequently emits luminescent light at 615 nm (Detected by microplate reader).
B) The human and SARS-CoV-2 proteins were expressed as eGFP and mCherry fusions and separated on SDS-PAGE gel (4–12% Tris-glycine) and visualized by in gel fluorescence scanning (Bio-RAD chemidoc MP). Last column: positive control (Figure 6B). M: marker, the proteins pair is each lane is annotated as shown at the end of this legend. The yield of protein production ranged between 10 nM and 60 nM for protein fusions. 1) N/RBM28. 2) ORF8/C2orf30. 3) ORF8/ERP44. 4) ORF8/PUSL1. 5) ORF6/MTCH1. 6) NSP2/RAP1GDS1. 7) NSP5/GPX1. 8) NSP7/MEL. 9) NSP7/QSOX2. 10) NSP7/RAB10. 11) NSP7/RHOA. 12) NSP7/SCCPDH. 13) NSP4/TIMM10. 14) NSP12/CRTC3. 15) NSP12/LARP4B. 16) NSP12/PPIL3. 17) NSP12/SLU7. 18) NSP14/IMDPH2. 19) ORF3a/ ARL6IP6. 20) E/ZC3H18. 21) S/ZDHHC5. 22) NSP13/GCC1. 23) NSP13/GCC2. 24) NSP13/GORASP1. 25) NSP13/PRKAR2A. 26) NSP13/HSBP1.

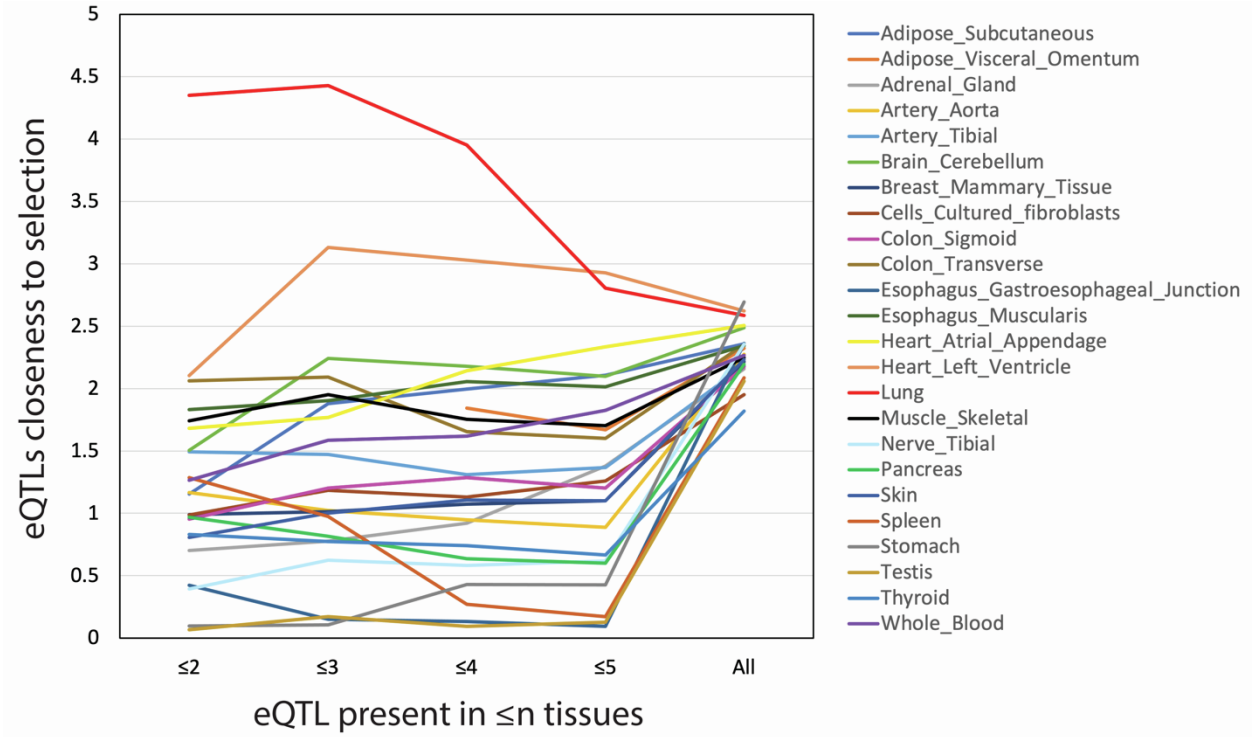


Figure S7. Proximity of iSAFE peaks to increasingly tissue-specific eQTLs. Related to STAR Methods and Figure 7.

The figure represents how much closer the eQTLs for the 42 CoV-VIPs selected 900 generations ago, are to the location of selection, compared to random expectations (Methods), as a function of their tissue specificity. The x axis represents the number of tissues where an eQTL was found by GTEx. For example for lung, ≤4 means that we tested the closeness to selection of lung eQTLs found in not more than three other tissues (four tissues in total). We did not include results with eQTLs found in only one tissue, because then many tissues did not have any, or very few eQTLs left. The y-axis represents the average (over tested eQTLs) difference between the expected log-distance, and the observed log-distance from the location of selection estimated by iSAFE (Methods; Figure S12). This difference is the numerator in the proximity ratio used in Figure 7, and the two should not be confused. Related to Figure 7.