Table S1. Total variance described with three components in Principle Component Analysis. **Exposed (E)** indicates the population exposed to chlorine dioxide, and **Non-Exposed (NE)** indicates the control population.

	Exposed (E)	**Non-Exposed (NE)**
Window size = 1	11.0%	11.2%
Window size = *M*/4	11.7%	13.2%
Window size = *M*/2	12.0%	13.4%

Table S2. Identification of mutation clusters with Pearson's correlation coefficients (*r*) from Zhong et al. 2017.

The *r* between any two mutations is calculated as following:

$$r = \frac{cov(s_1, s_2)}{\sigma_{s1}\sigma_{s2}} = \frac{\frac{1}{N}\sum(s_{1i} - \bar{s}_1)(s_{2i} - \bar{s}_2)}{\sigma_{s1}\sigma_{s2}}$$

where *N* is the number of virus populations, s_i is the increase in frequency of the mutation compared to WT for each virus population, $\bar{s}$ is the mean increase in frequency of the mutation among all the populations and σ is the standard deviation of the increase in frequency among all the populations. $r > 0.95$ is assumed to indicate correlated emergence of mutations.

I	II	III	IV	V	VI	VII
3162	2835	3552	1761	2937	C6006T	6745
4384	2844	5389	6989	3101	C6562T	7383
4552	2849	5650				
5203	2850	6586				
5788	3162	6976				
5893	3170					
T6006C	3233					
6061						
6409						
T6562C						

Table S3. Hierarchical clustering of the distributed vectors of the alleles. **Exposed (E)** indicates the population exposed to chlorine dioxide, and **Non-Exposed (NE)** indicates the control population.

	***w*=1**		***w*=*M*/4**		***w*=*M*/2**	
Cluster	**Exposed**	**Non-exposed**	**Exposed**	**Non-exposed**	**Exposed**	**Non-exposed**
I	5323	1761	7383	6140	6112	6191
	6006	7282	5203	3139	6061	2521
	5650	1210	2849	1834	724	5710
	2835	5506	2850	2521	2850	4519
	6061	5211	5237	4454	4552	7247
	4552	411	7247	5821	6745	949
	3170	2497	6061	5585	7249	1082
	3162		2283	3644	5203	6649
	3233		7244	7249	2283	7144
	6586		4384	7250	6562	7246
	2632		6006			1866
	6976		4687			4454
	1761		5783			3644
	5389		6562			5585
II	2854	4471	5964	7244	3170	3139
	6190	5083	7250	838	7246	6403
	3103	4148	1660	2851	5818	4148
	7240	5821	1761	5506	2835	7359
	5634	6403	5893	5323	6006	
	7250	1243	6745	3271	7383	
	2844	6085	2632	1437	1666	
	2849	3644	4552	4666	5893	
	5203	4579	2844		5323	
	7383	7347	6409		2844	

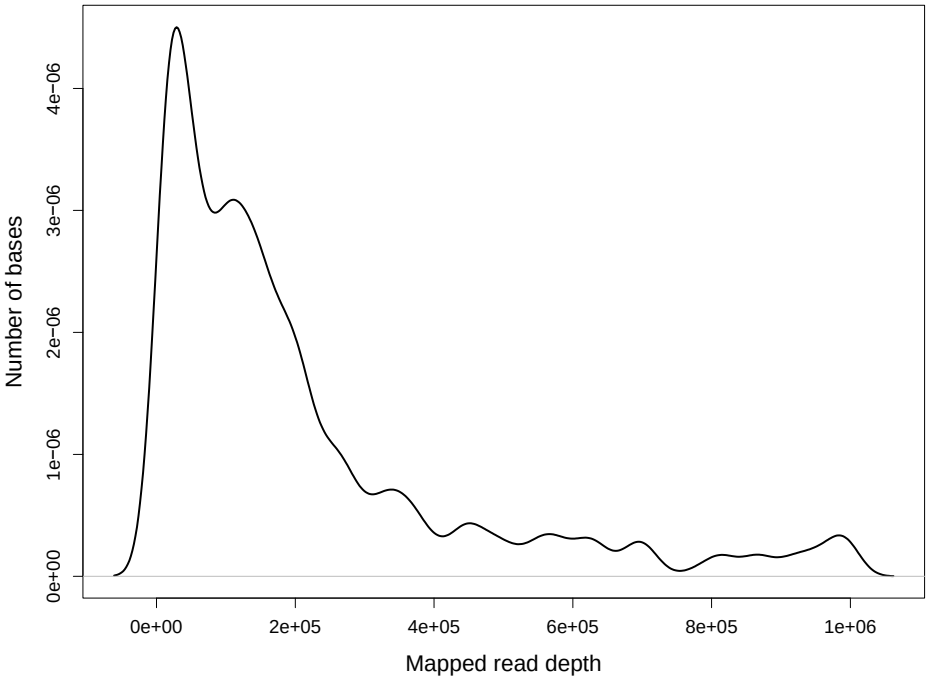
		3139 7069	3233 3170 3103 2835 3162		3233 2849 5788 3103 4384	
III	5813 7244 2850 7249 5023 5927 7388 3503 5643 5641 5628 5961	1672 7250 1450 6284 6649 1831 3841 659 7359 7024 1437 5306	4255 5572 7249 5963 5323 7246 724 1666	2497 3697 659 1450 4471 6190 7024	1243 2854 5965 7388 5023 5389 5820 6871 7242 4053 5628 5636 5817	5211 7054 411 5506 7282 3697 7250 6940 7244
IV	1243 3861 5965 1666 5638 5820 5960 4915 7359 726 5959	1082 7240 4546 5323 5585 2937 6989 7247	5962 5965 5389 6191 6924 6112 5634 6586	7290 3411 7054 6284 6191 7247 7347 7359 4148 7346	5650 6976 7240 4255 3518 7244 1761 5237 2632 3162 6409	6085 1437 2851 1210 1243 1450 2497 5109 1831 6190 3841

	724 4053				7247	4666 3271 5083
V	4766 1480 5636 5817 5819 5632 5633 7247 5963 7255 5962 6191 4255 7246 5631 1660 6924	5710 6991 3101 4519 4666 1008 3367 4454 3285 2521 7346 7246 7249	5960 6871 5638 2863 3503 3861 4766 5636 726 6190 4915 5961	1210 1831 3101 3841 4546 6991 2937 6085 5306 6989 1243 1761	6191 6469 3861 7359 5814 2863 5632 5958 3552 5634 726 4915 5964 5631 7255 5963 5633 5819	1008 3101 2937 6989 6991 659 4546 5306 1761 5323
VI	5958 3518 5814 2863 6871 6112 6469	3271 3585 838 949 2851 1866 7290	5817 7383 6469 5958 7255 5820 3518 5633	1082 1672 1866 7282 411 3285 6649 7069	5957 5959 1480 6190 4766 5641 3503 5643	3285 7024 4471 7249 7346 7069 7347 7290

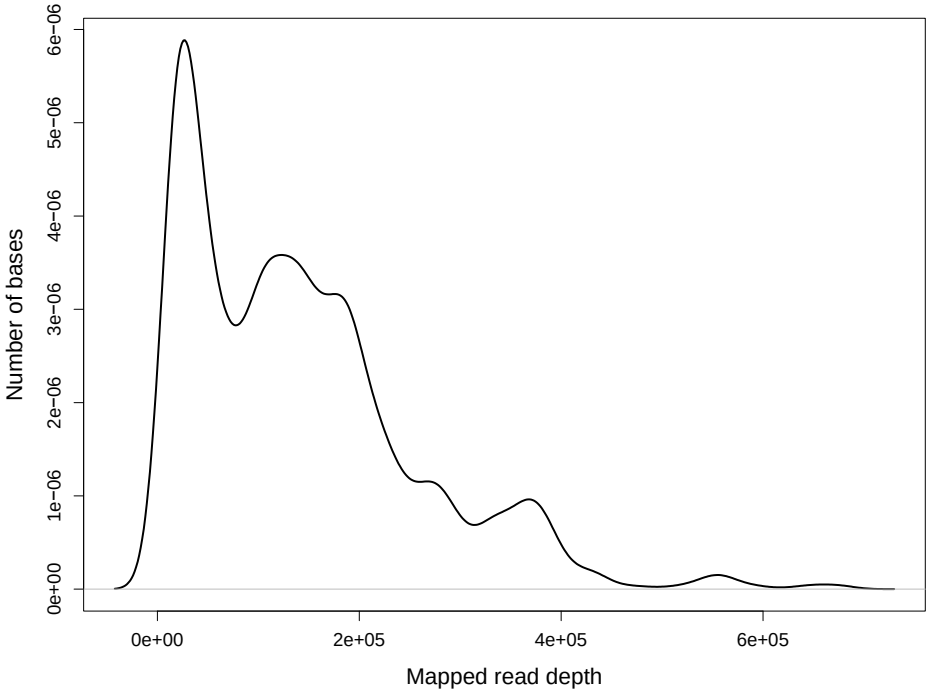
			5632 5641 5643 5628 5814 5819	7242 3367 5083	5638 6924	838 7240
VII	6745 3552 7242 2283 5572 5942 4384 4687 5788 5893 6562 5237 6409	7242 7244 5467 6190 6940 7286 3411 3697 6191 7144 7054 1834 5109	2854 5650 7240 7242 5631 5023 6976 4053 5959 5957 1480 3552 1243 5818 7359	3585 6403 7286 5710 5467 5109 5211 4519 949 7144 1008 7240 4579 7246	1660 6586 5960 5961 5962 7250 4687 5572	1834 3367 4579 1672 3585 5467 7286 3411 7242 5821 6284

Figure S1. Coverage from Illumina HiSeq 2500 of read depth histograms (A) EA, (B) EB, (C) NEA, (D) NEB and of genome-wide read depths (E) EA, (F) EB, (G) NEA, (H) NEB. As the datasets were pool-sequenced from the short reads, the difference in the raw reads of each allele between two sampling time points was reconstructed by simulating each site in the virus genome as the binomial distribution.

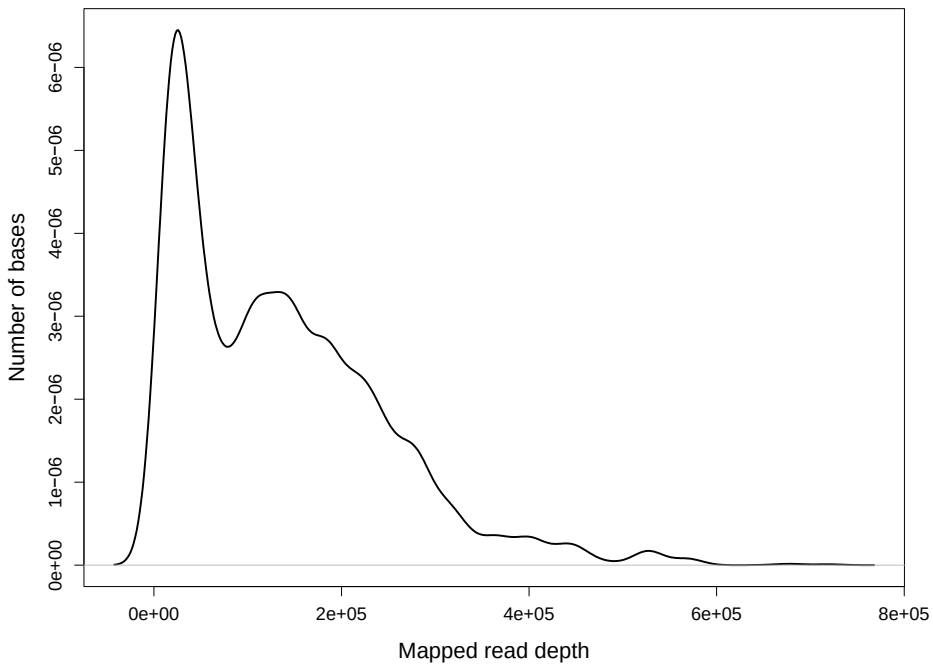
A



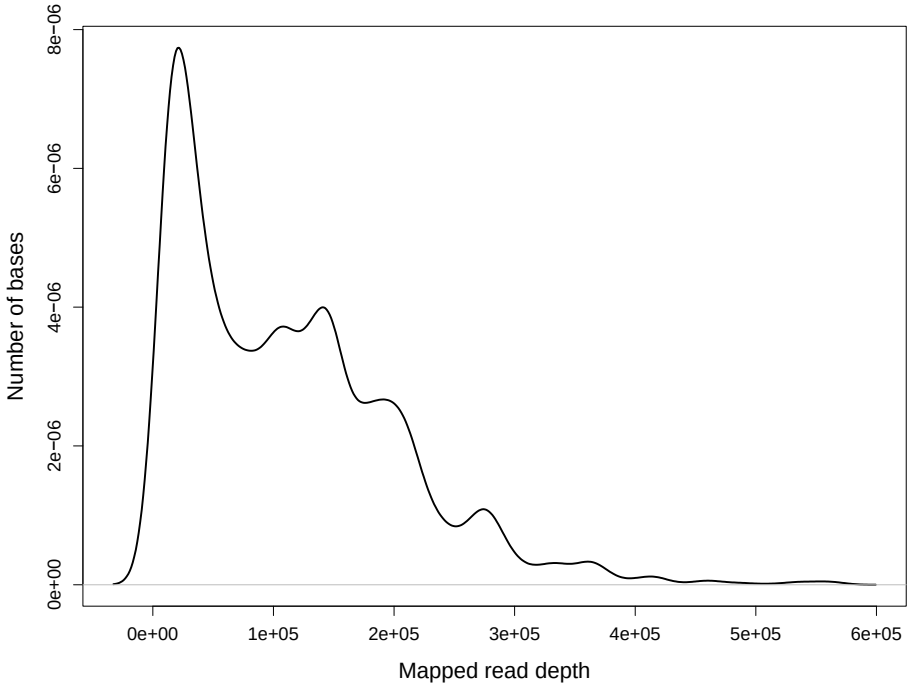
B



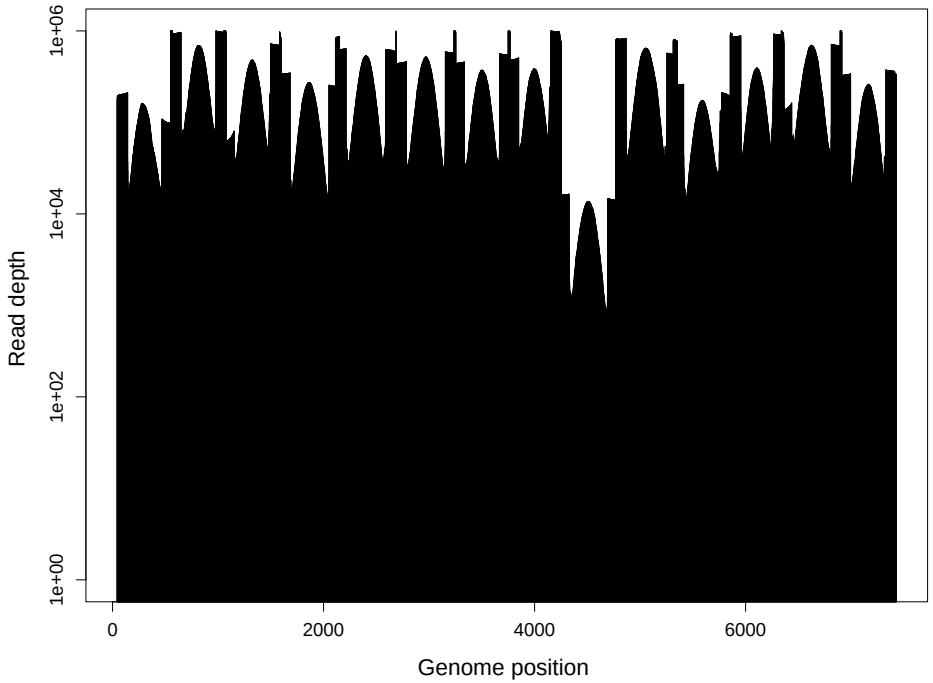
C


Number of bases
0e+00
1e−06
2e−06
3e−06
4e−06
5e−06
6e−06
0e+00
2e+05
4e+05
6e+05
8e+05
Mapped read depth

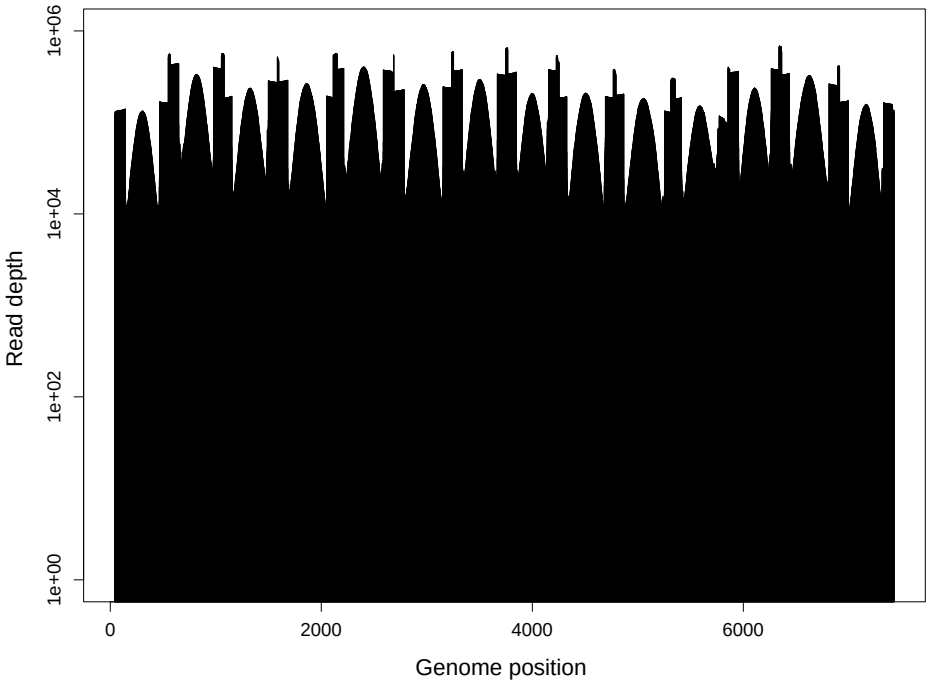
D


Number of bases
0e+00
2e−06
4e−06
6e−06
8e−06
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
6e+05
Mapped read depth

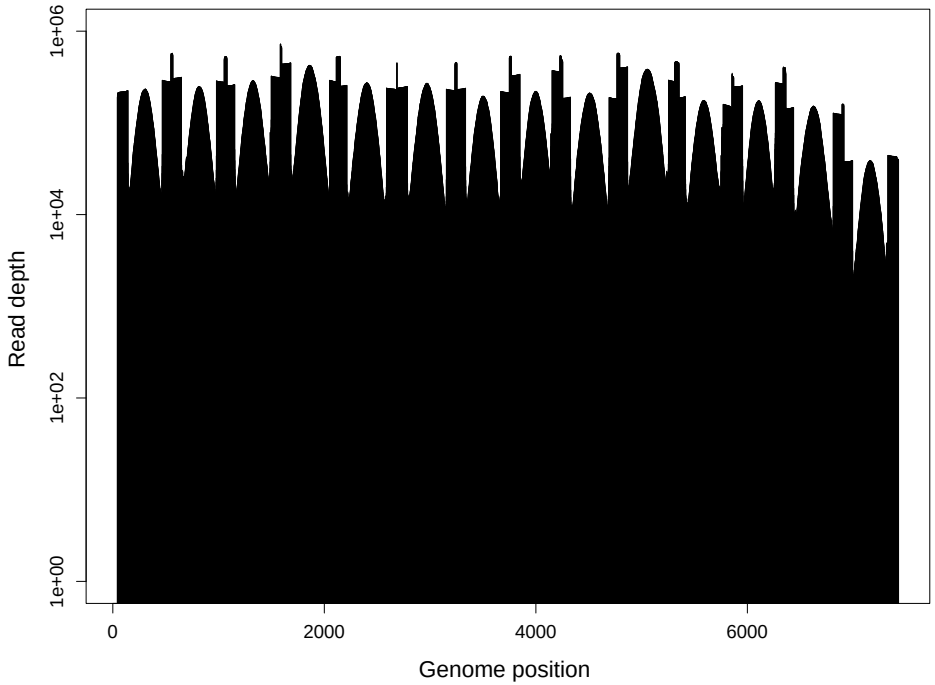
E


Read depth
1e+06
1e+04
1e+02
1e+00
0
2000
4000
6000
Genome position

F


Read depth
1e+06
1e+04
1e+02
1e+00
0
2000
4000
6000
Genome position

G


Read depth
1e+00
1e+02
1e+04
1e+06
0
2000
4000
6000
Genome position

H

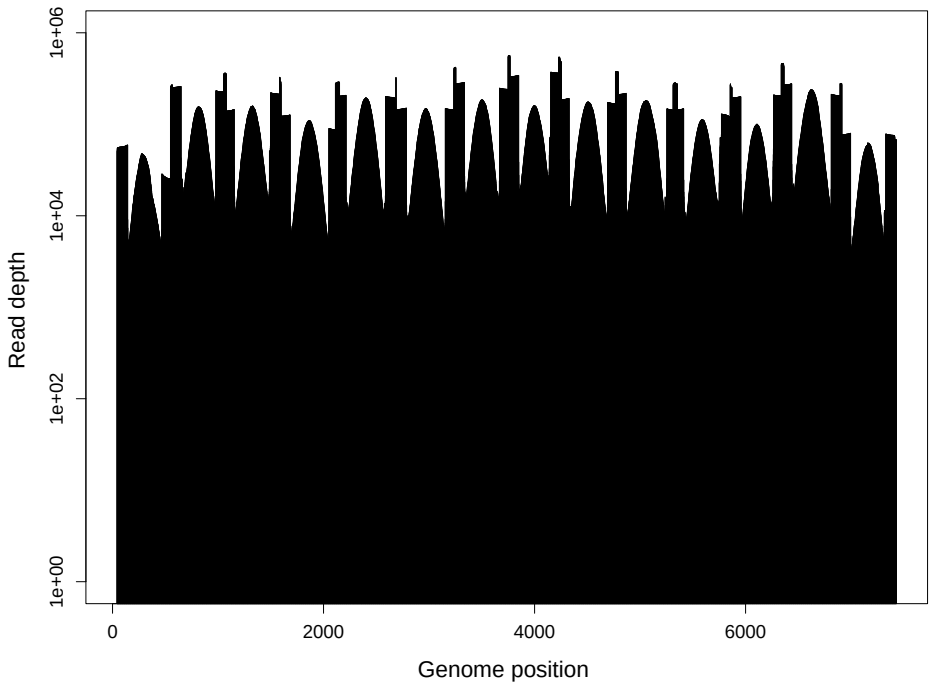

Read depth
1e+00
1e+02
1e+04
1e+06
0
2000
4000
6000
Genome position

Figure S2. Computational graph of training the nucleotide skip-gram neural network from TensorBoard; nodes represent operations, solid lines represent data flow, and dotted lines represent control dependence edges.

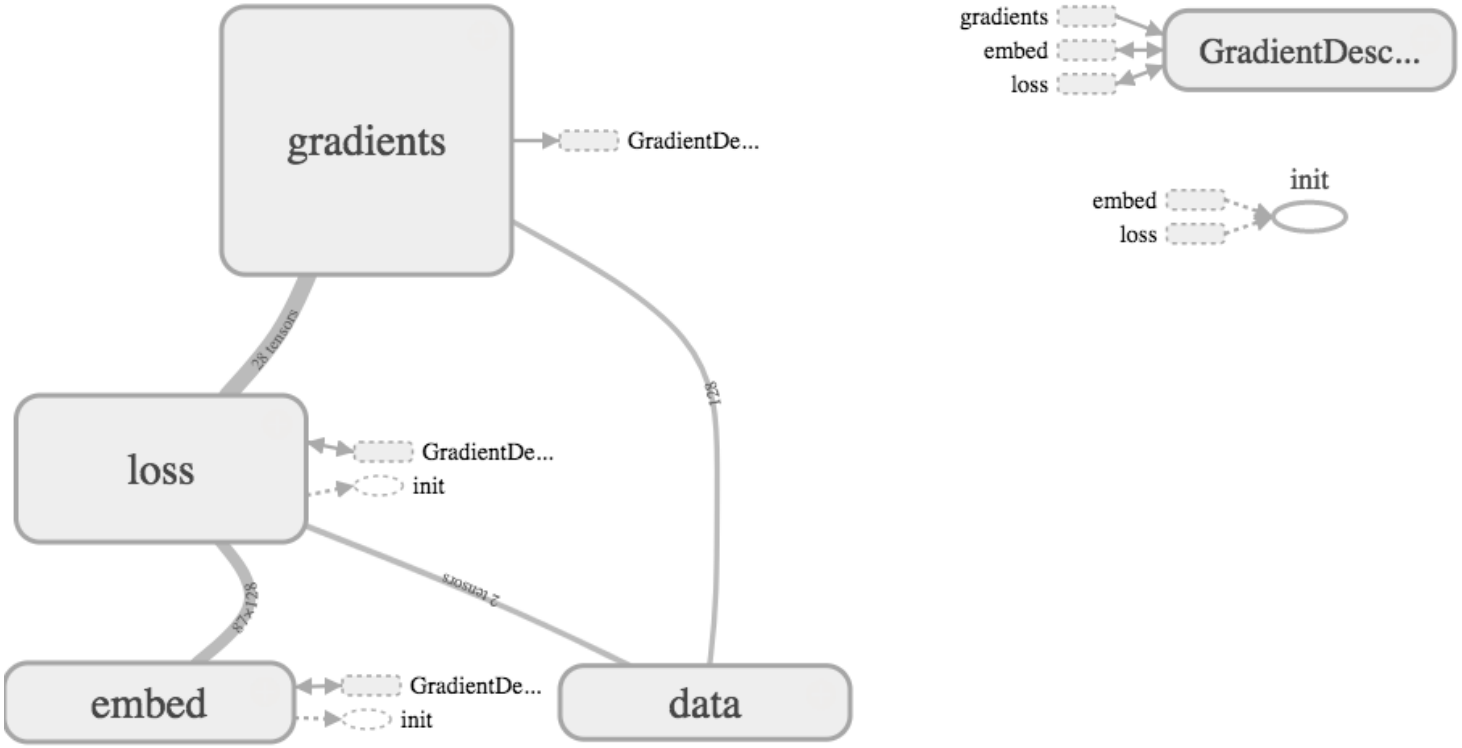


Figure S3. Workflow schematic of training the nucleotide skip-gram neural network in TensorFlow. For feature learning, the neural network is trained in TensorFlow (Version 1.2.1) to optimize the probability for every allele in the datasets of being the nearby allele given the centre allele.

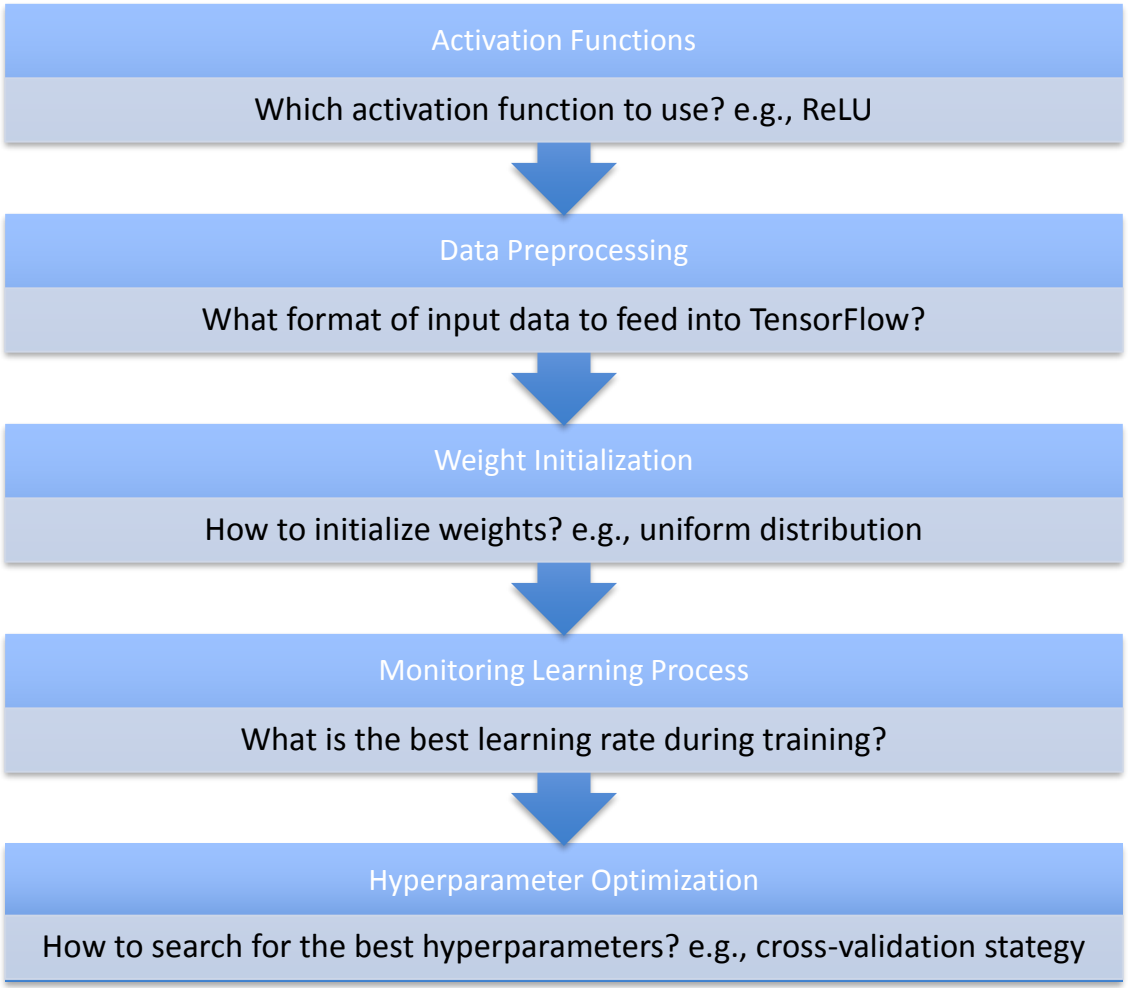
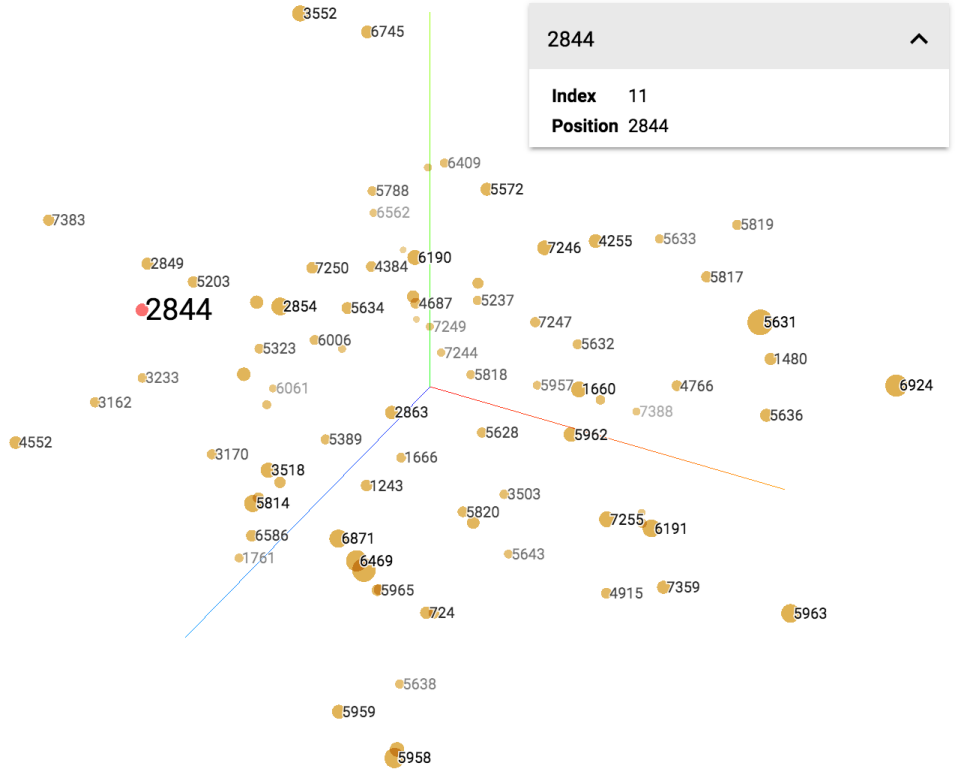


Figure S4. Cosine distances of the allele embeddings in three-dimensional space from TensorBoard with the window size as *w*=1: A. Exposed and B. Non-Exposed. Each point is indexed to the nucleotide position in the genome, and the allele of interest in the exposed population (P129Q denoted as Position 2844) and the non-exposed population (H215N denoted as Position 3101) is highlighted, in order to indicate candidate mutations of adaptation.

A



B

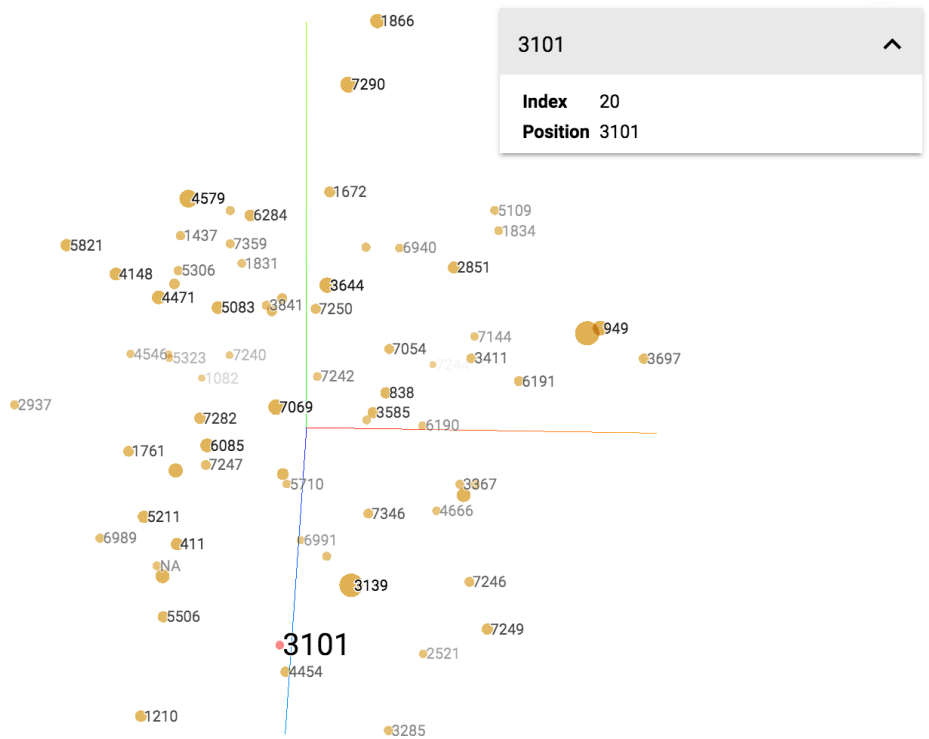
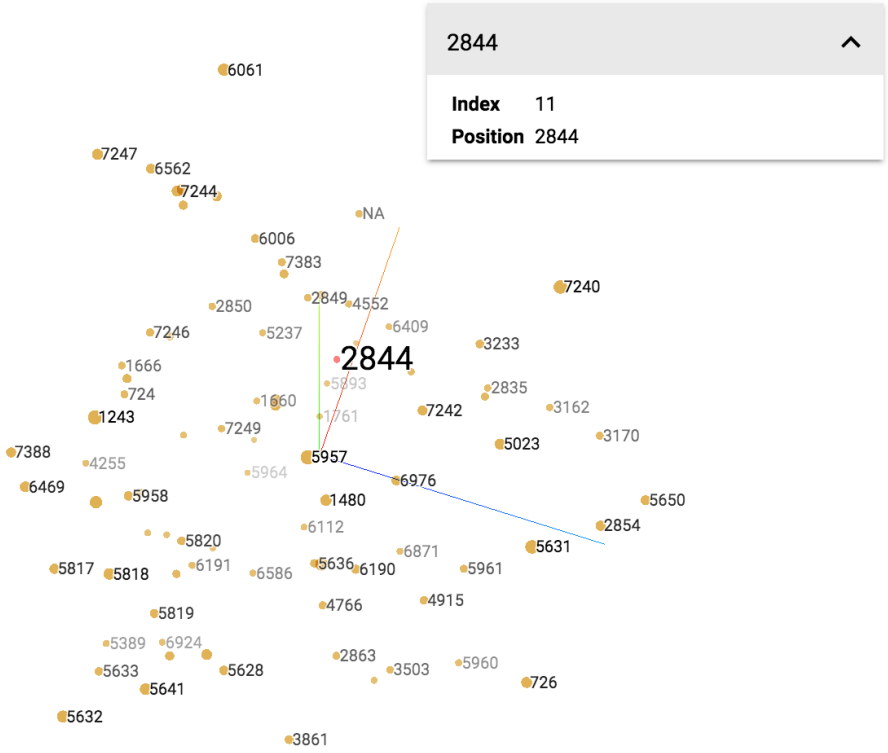


Figure S5. Cosine distances of the allele embeddings in three-dimensional space from TensorBoard with the window size as *w*=*M*/4: A. Exposed and B. Non-Exposed. Each point is indexed to the nucleotide position in the genome, and the allele of interest in the exposed population (P129Q denoted as Position 2844) and the non-exposed population (H215N denoted as Position 3101) is highlighted, in order to indicate candidate mutations of adaptation.

A



B

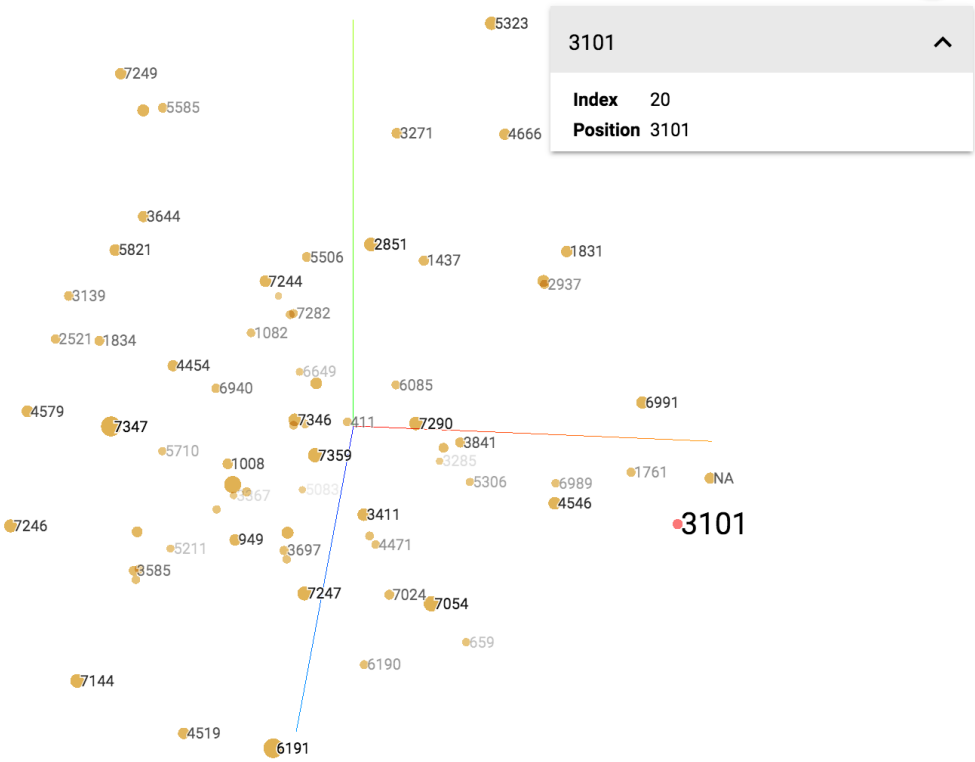
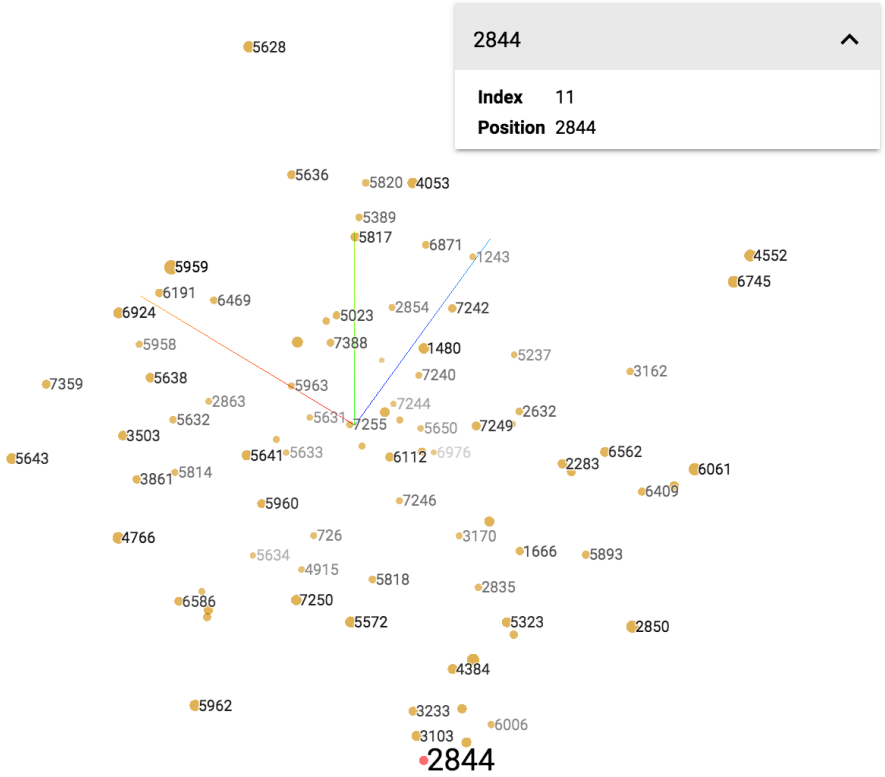


Figure S6. Cosine distances of the allele embeddings in three-dimensional space from TensorBoard with the window size as *w=M/*2: A. Exposed and B. Non-Exposed. Each point is indexed to the nucleotide position in the genome, and the allele of interest in the exposed population (P129Q denoted as Position 2844) and the non-exposed population (H215N denoted as Position 3101) is highlighted, in order to indicate candidate mutations of adaptation.

A



B

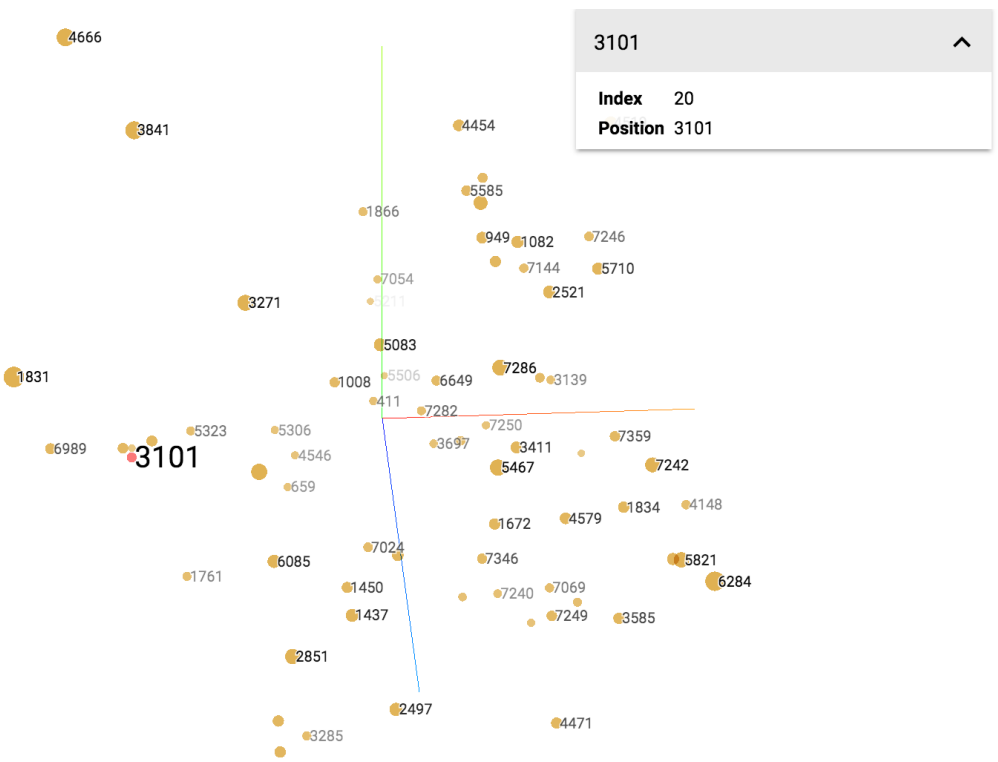
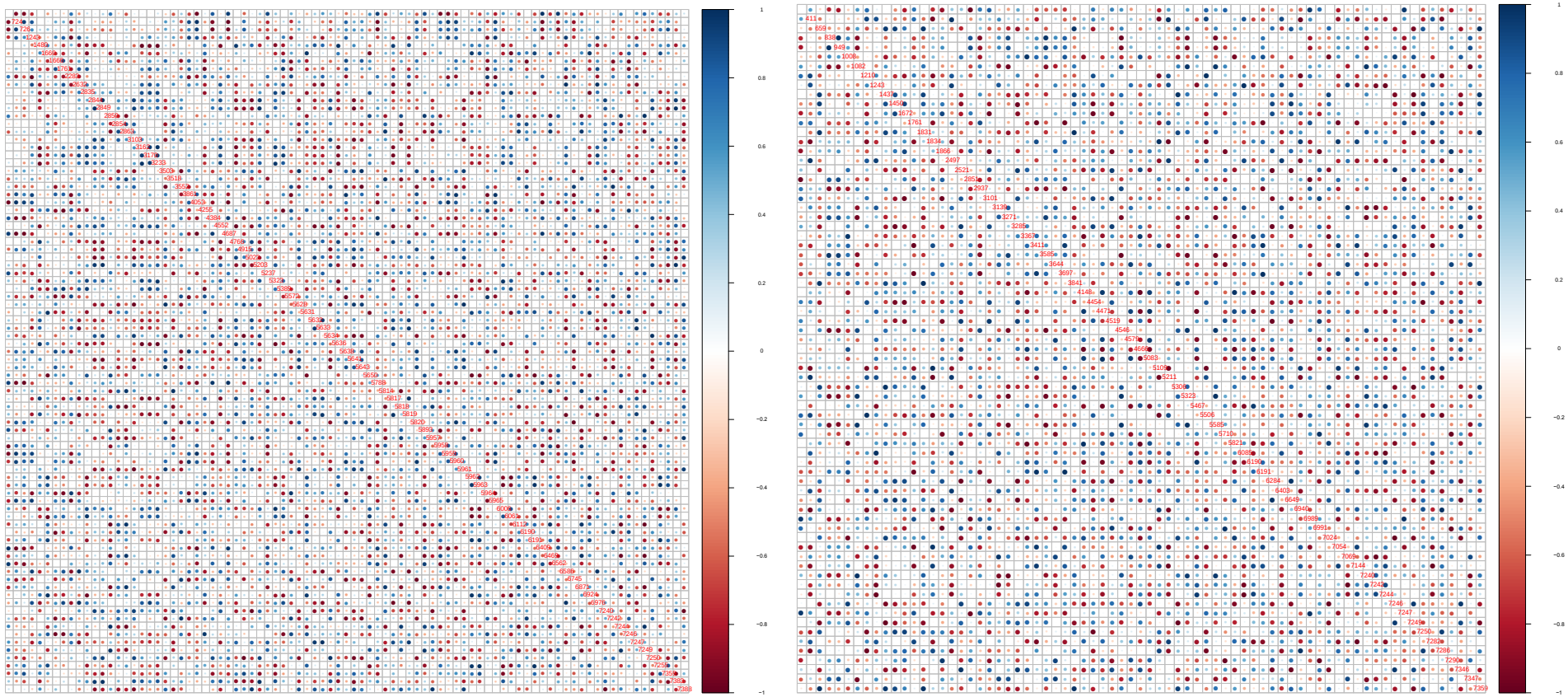
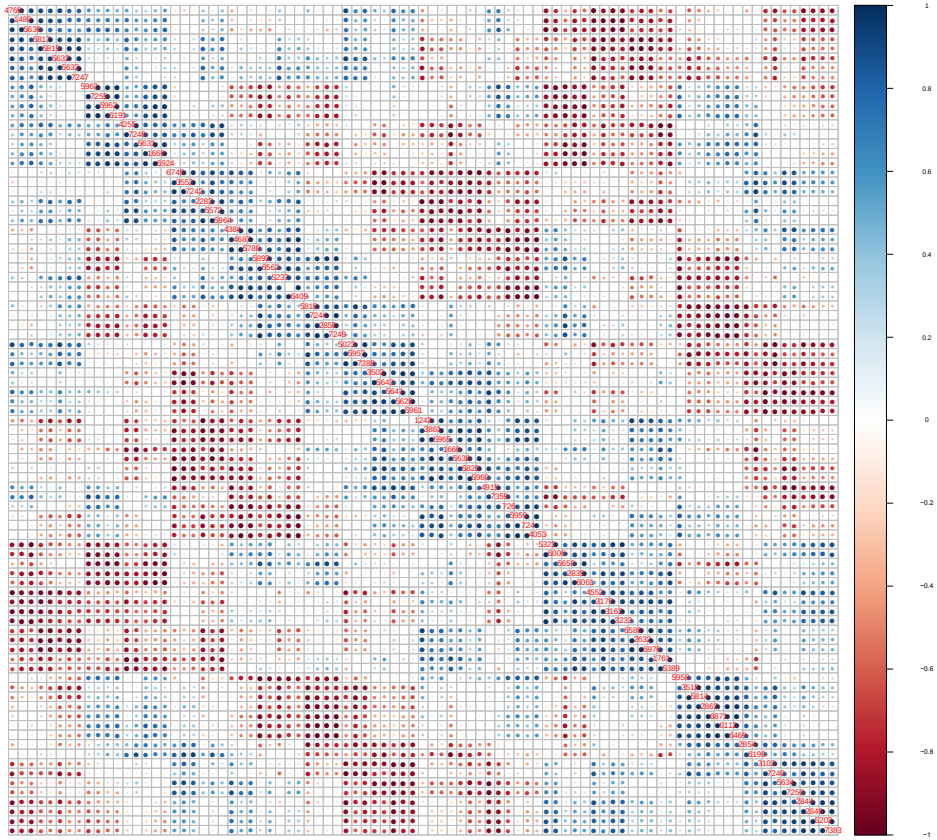


Figure S7. Pairwise correlation map of the first three PCA components of the allele embeddings from TensorBoard with *w*=1. The alleles are arranged in the genomic order: A. Exposed and B. Non-Exposed, and in the hierarchical clustering: C. Exposed and D. Non-Exposed. Positive correlations are in blue and negative correlations are in red, with the color intensity proportional to the correlation coefficients ranging from 1 to -1.

A **B**



C



D

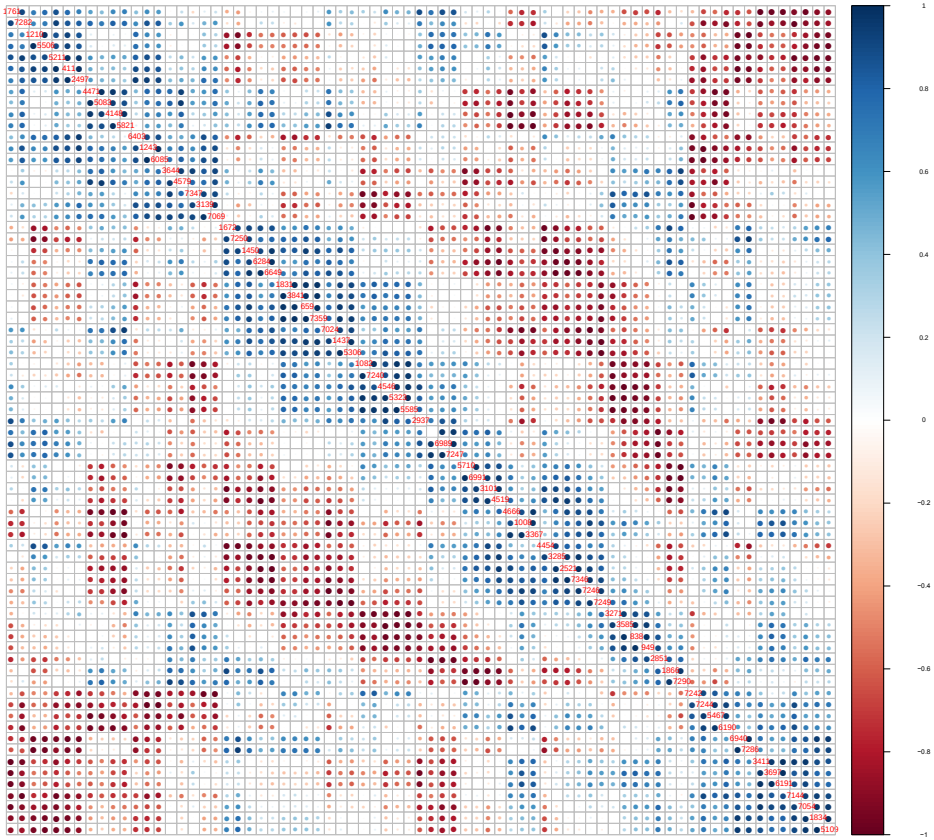
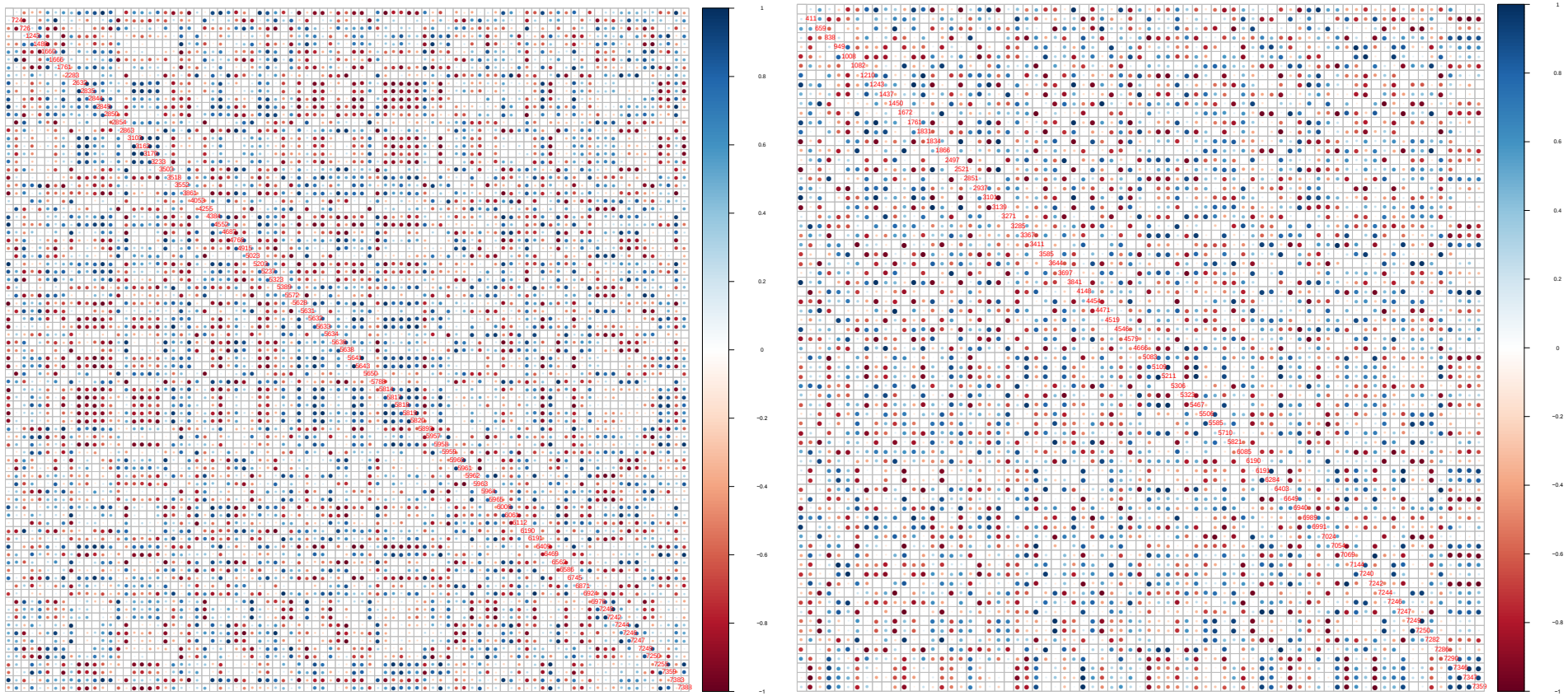
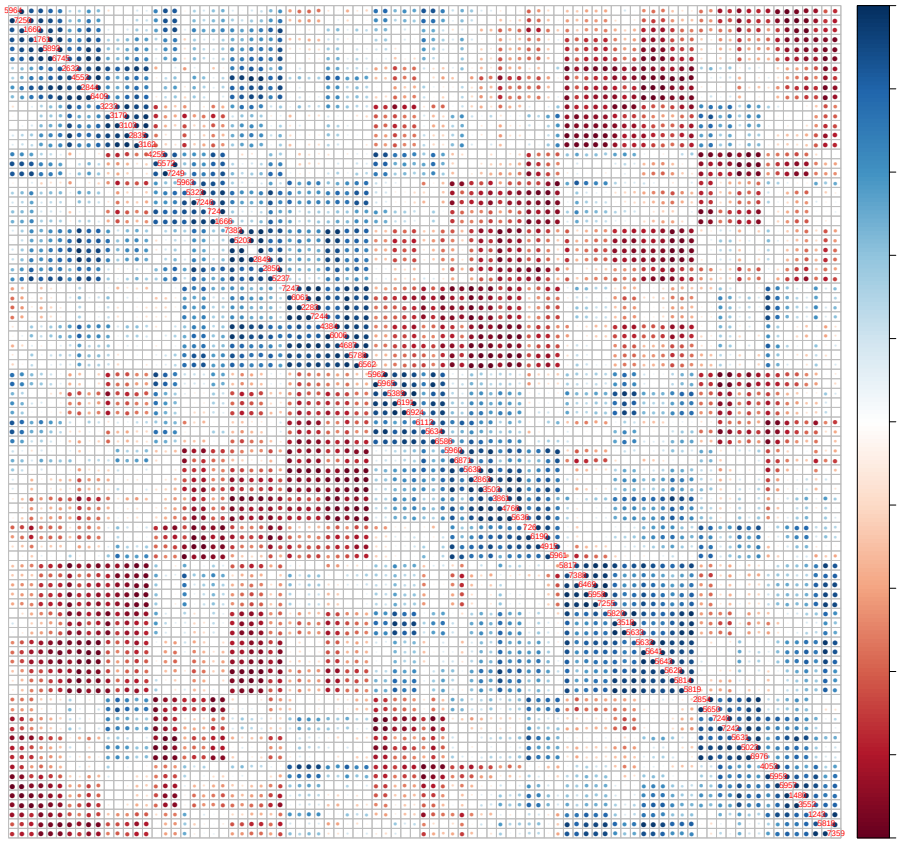


Figure S8. Pairwise correlation map of the first three PCA components of the allele embeddings from TensorBoard with *w*=*M*/4. The alleles are arranged in the genomic order: A. Exposed and B. Non-Exposed, and in the hierarchical clustering: C. Exposed and D. Non-Exposed. Positive correlations are in blue and negative correlations are in red, with the color intensity proportional to the correlation coefficients ranging from 1 to -1.

A **B**



C



D

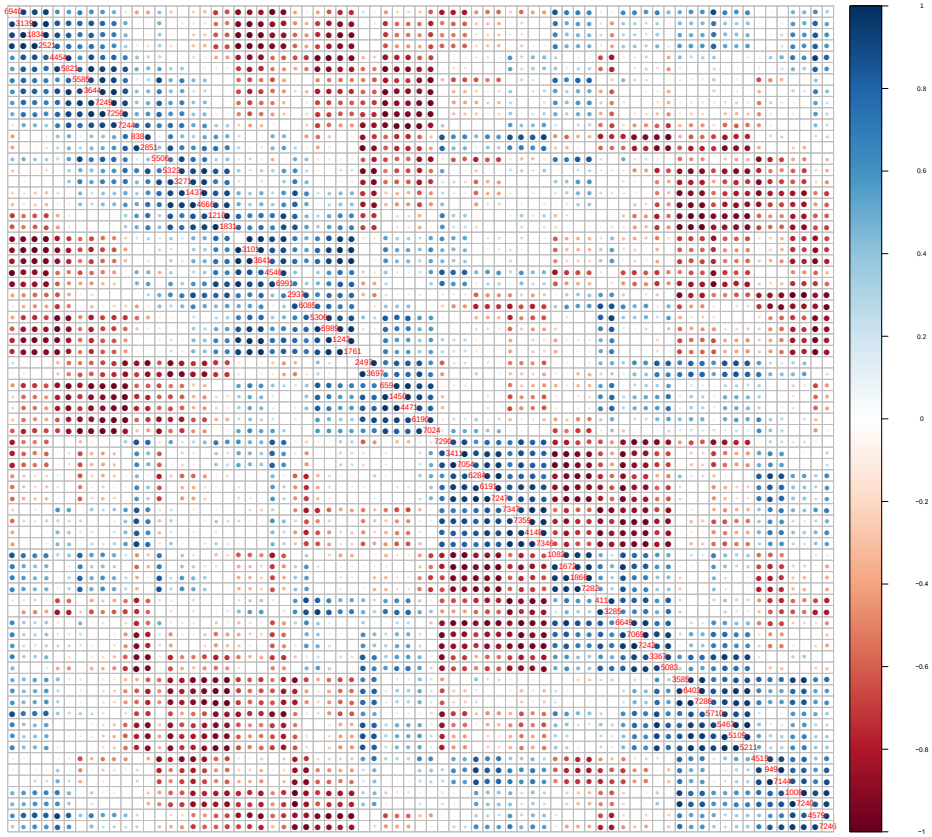
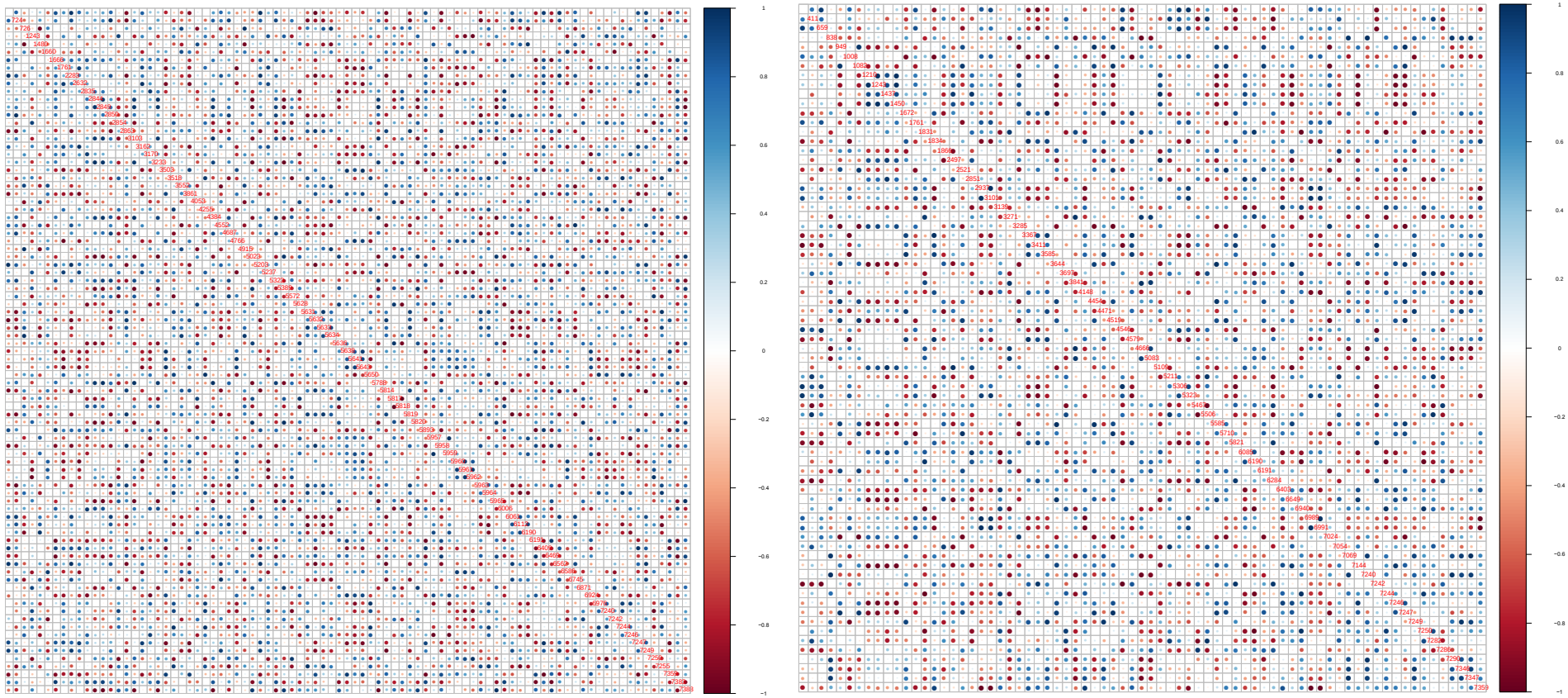
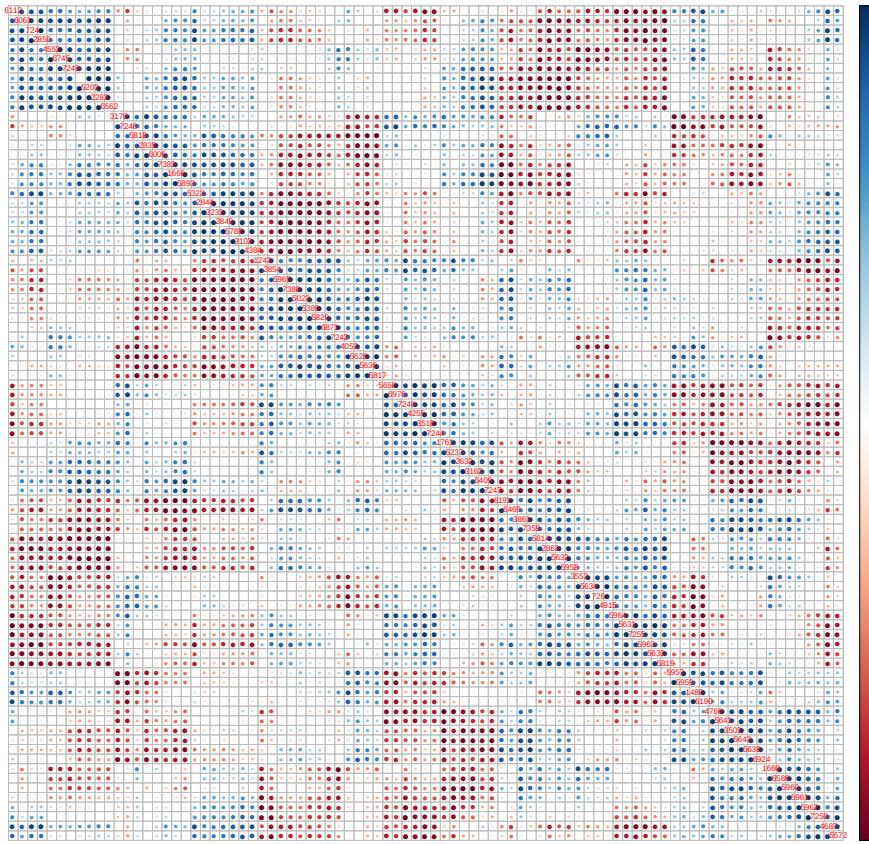


Figure S9. Pairwise correlation map of the first three PCA components of the allele embeddings from TensorBoard with *w*=*M*/2. The alleles are arranged in the genomic order: A. Exposed and B. Non-Exposed, and in the hierarchical clustering: C. Exposed and D. Non-Exposed. Positive correlations are in blue and negative correlations are in red, with the color intensity proportional to the correlation coefficients ranging from 1 to -1.

A **B**



C



D

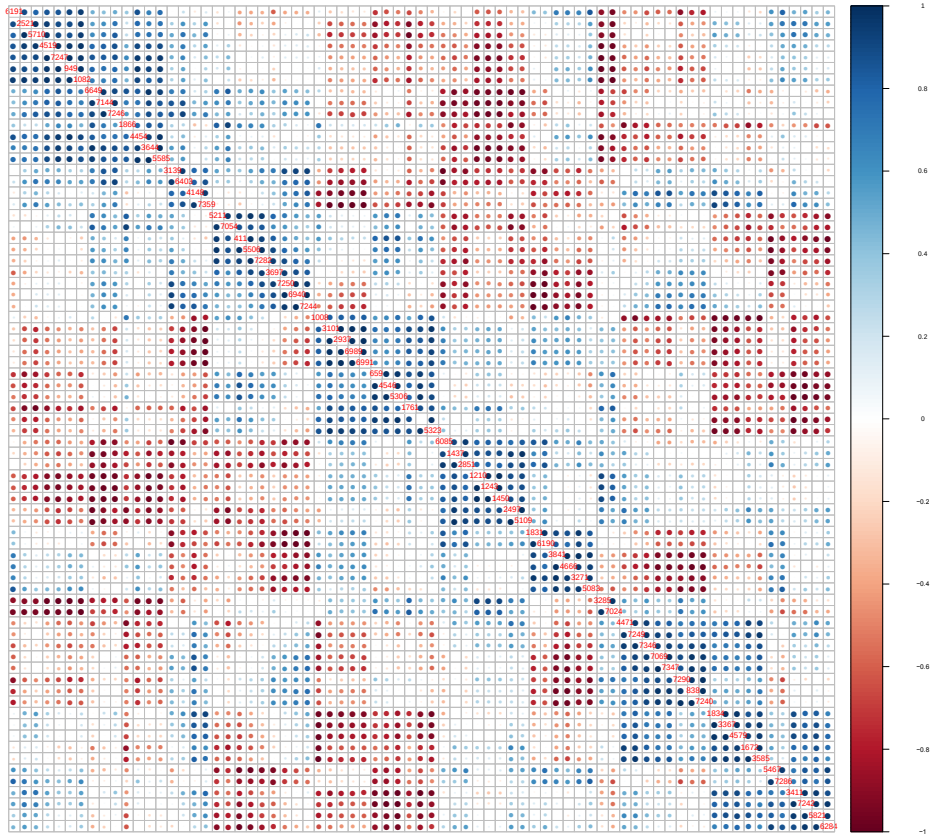
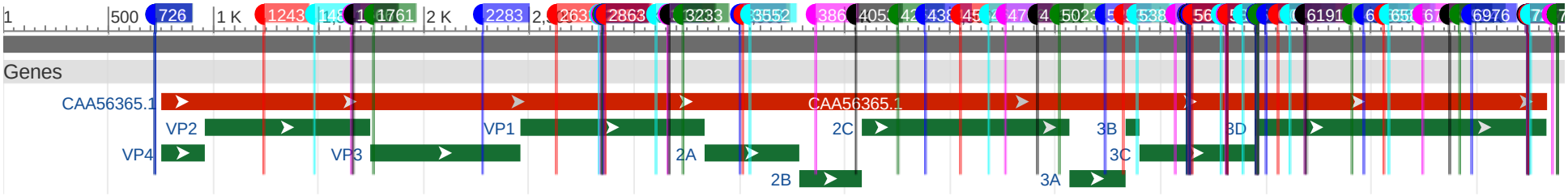


Figure S10. Distribution of the alleles along the echovirus genome: A. Exposed and B. Non-Exposed. The visualization reveals the alleles are evenly distributed along the genome, making the assumption of the window of genetic interactions as a recombination range biologically plausible. The images were generated by NCBI Graphics based on Echo Virus 11 genomic DNA (GenBank: X80059.1).

A



B