

Supplementary Figure Legends

Supplementary Figure S1: UMAP visualisation of full dataset coloured by DBSCAN clusters. Points which are unassigned to a cluster (Cluster 0) are coloured in gray.

Supplementary Figure S2: UMAP visualisation of full dataset coloured by CDC clade.

Supplementary Figure S3: UMAP visualisation of full dataset coloured by country.

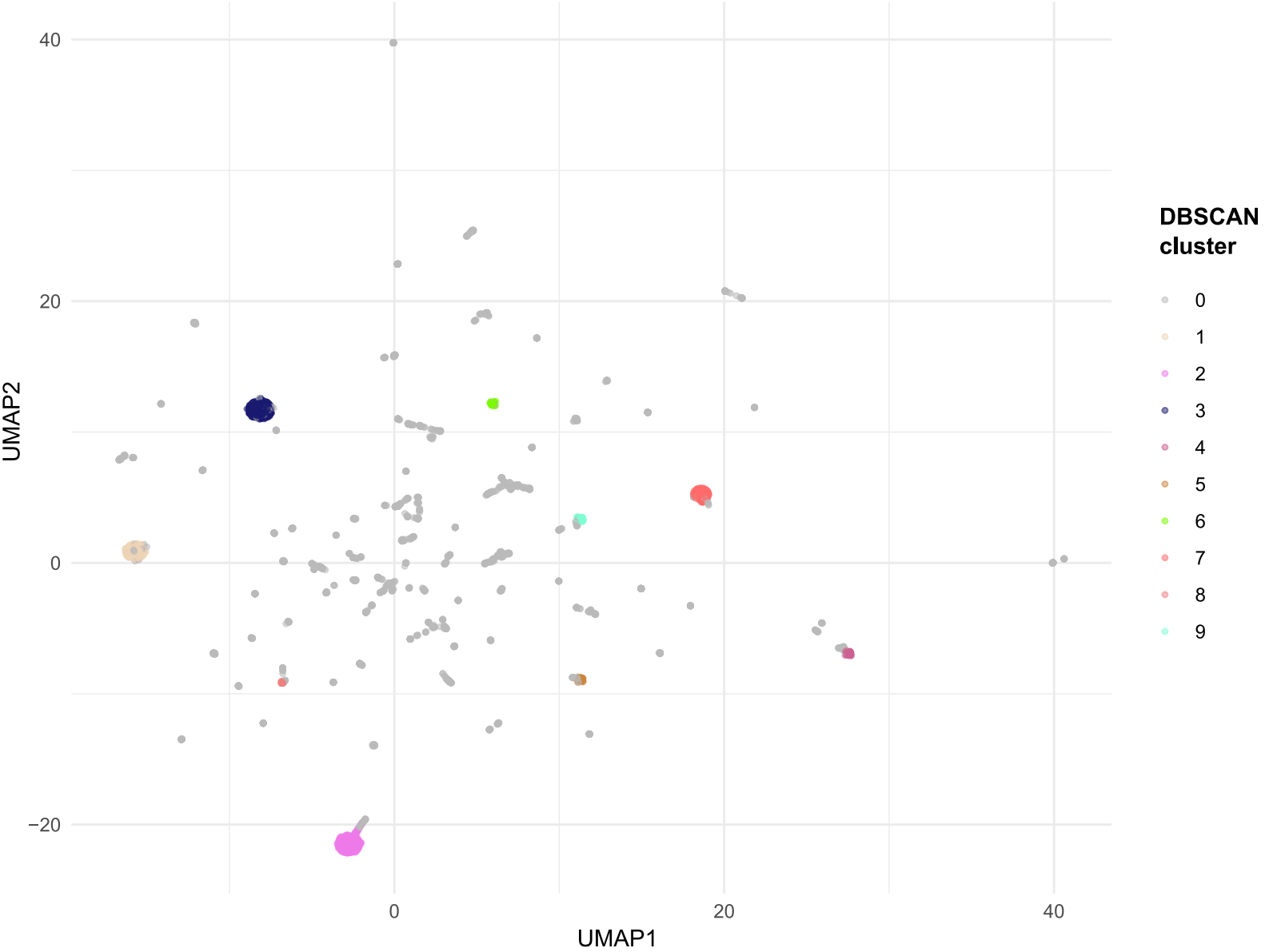
Supplementary Figure S4: Comparative Manhattan plots showing $p_{-\log_{10}}$ values for logistic regression results and importance scores for VariantSpark results. LR; Firth's logistic regression, VS; VariantSpark

Supplementary Figure S5: Association metrics for 4-SNV combination 4-SNV-1 (6319:21801:22346:25563). 0 denotes reference allele, and 1 denotes the alternate allele for each 4-SNV locus. Rate refers to the ratio of controls:cases in that particular allele combination subset, whereas the purity refers to the purity of the control pool of samples in the subset. Set rate and set purity denote the ratio of controls:cases and purity of controls for the whole dataset.

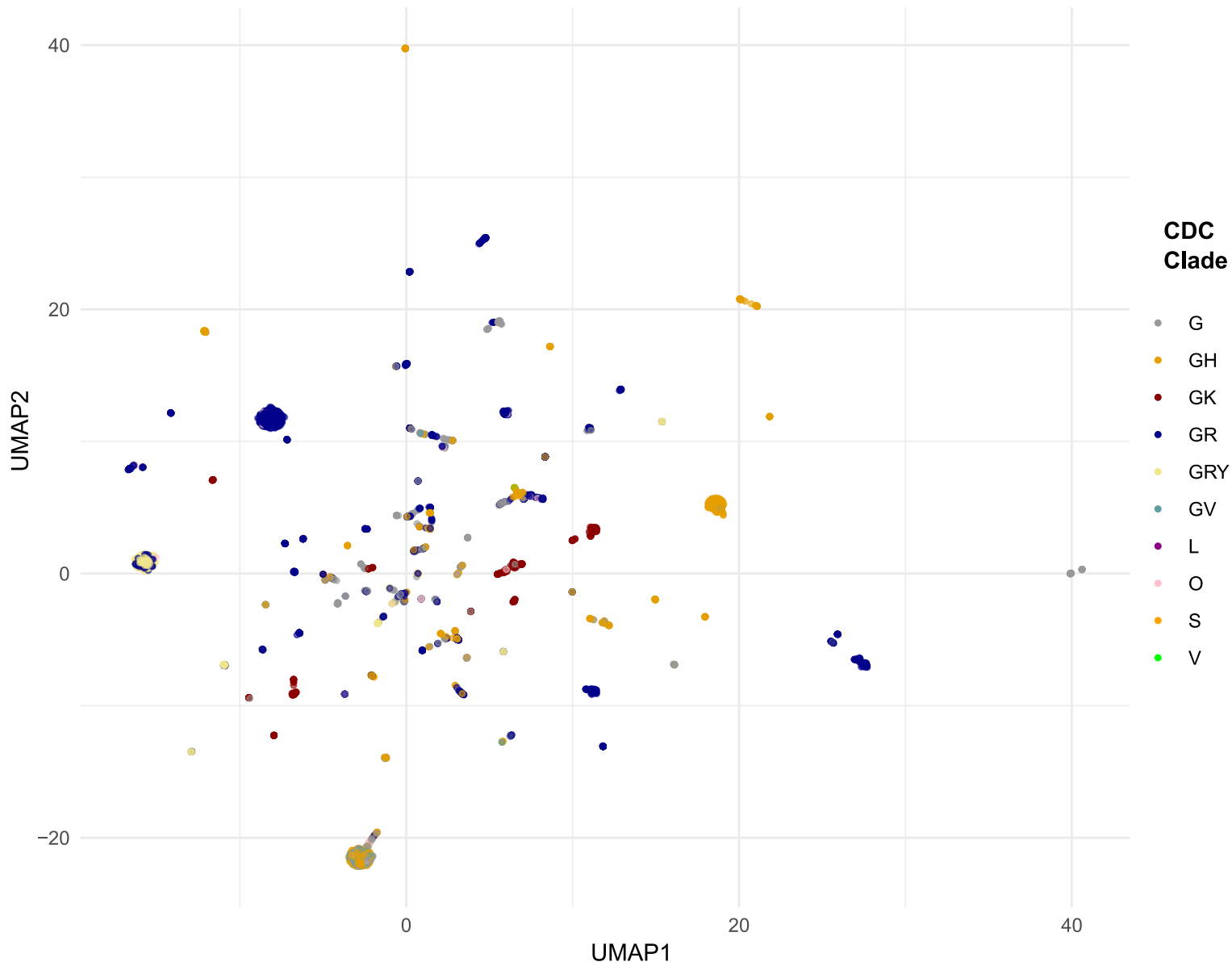
Supplementary Figure S6: Association metrics for 4-SNV combination 4-SNV-2 (19288:22343:26149:29144). 0 denotes reference allele, and 1 denotes the alternate allele for each 4-SNV locus. Rate refers to the ratio of controls:cases in that particular allele combination subset, whereas the purity refers to the purity of the control pool of samples in the subset. Set rate and set purity denote the ratio of controls:cases and purity of controls for the whole dataset.

Supplementary Figure S7: Association metrics for 4-SNV combination 4-SNV-3 (6319:19288:22343:29144). 0 denotes reference allele, and 1 denotes the alternate allele for each 4-SNV locus. Rate refers to the ratio of controls:cases in that particular allele combination subset, whereas the purity refers to the purity of the control pool of samples in the subset. Set rate and set purity denote the ratio of controls:cases and purity of controls for the whole dataset.

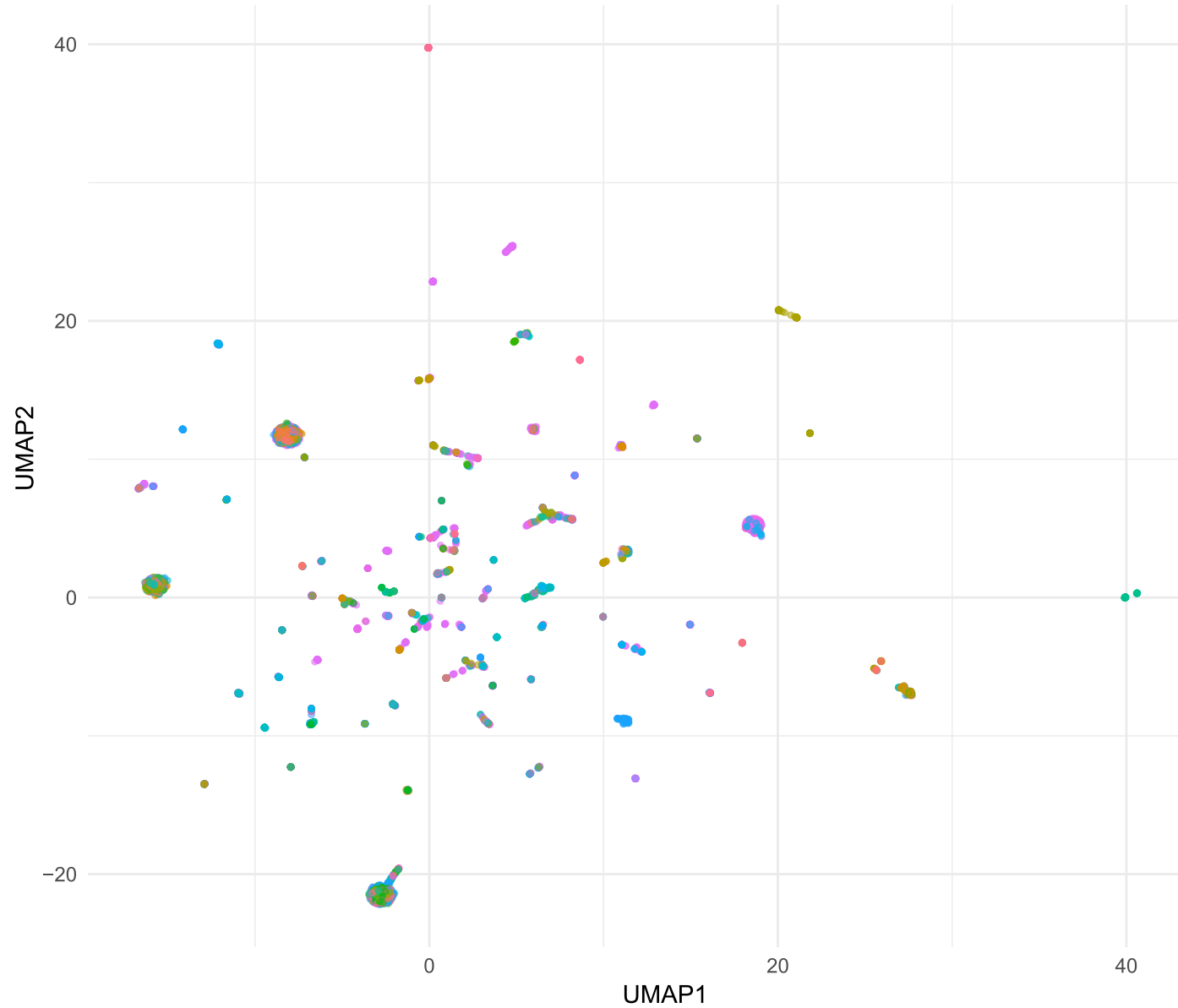
Supplementary Figure S1



Supplementary Figure S2



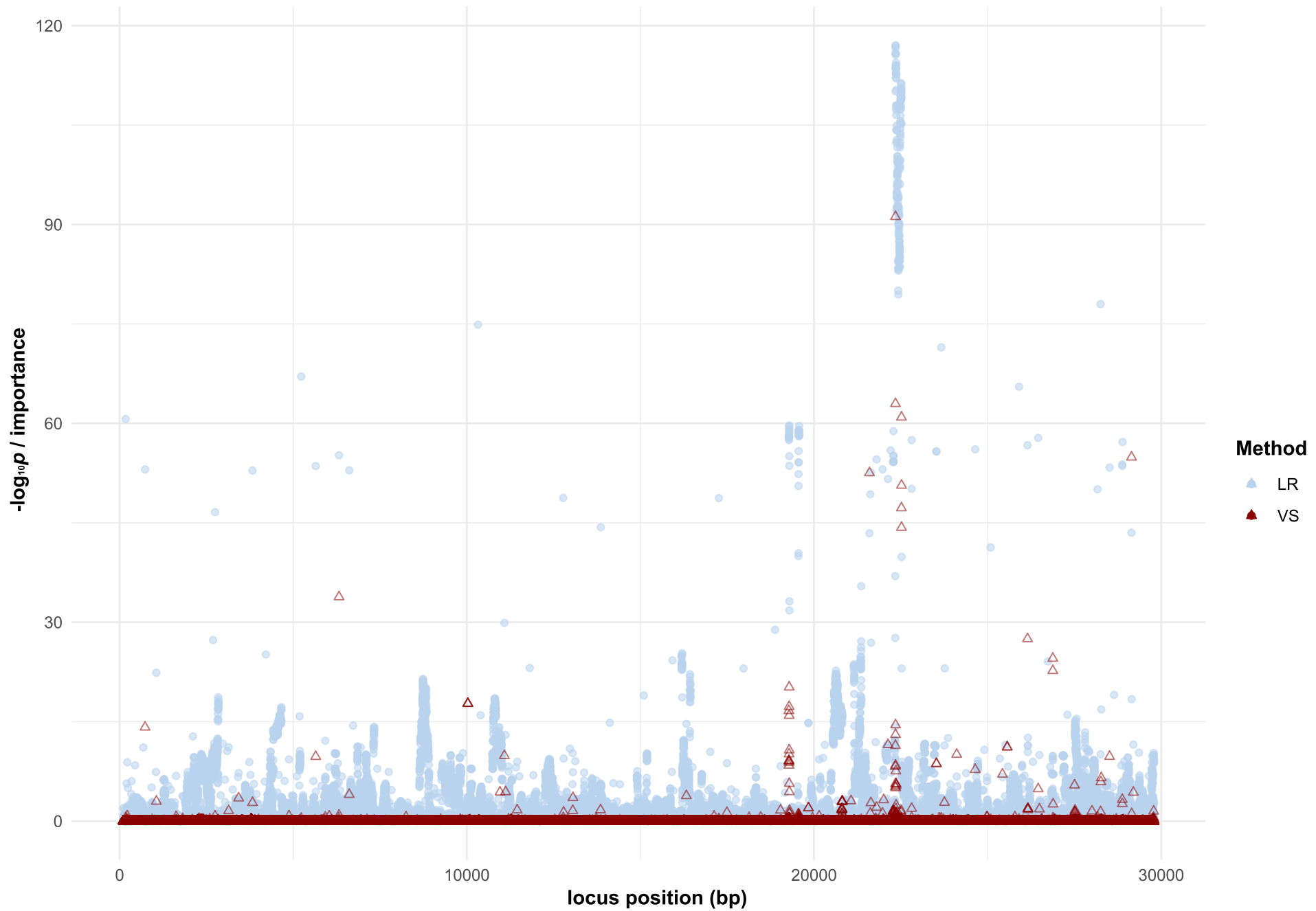
Supplementary Figure S3



Country

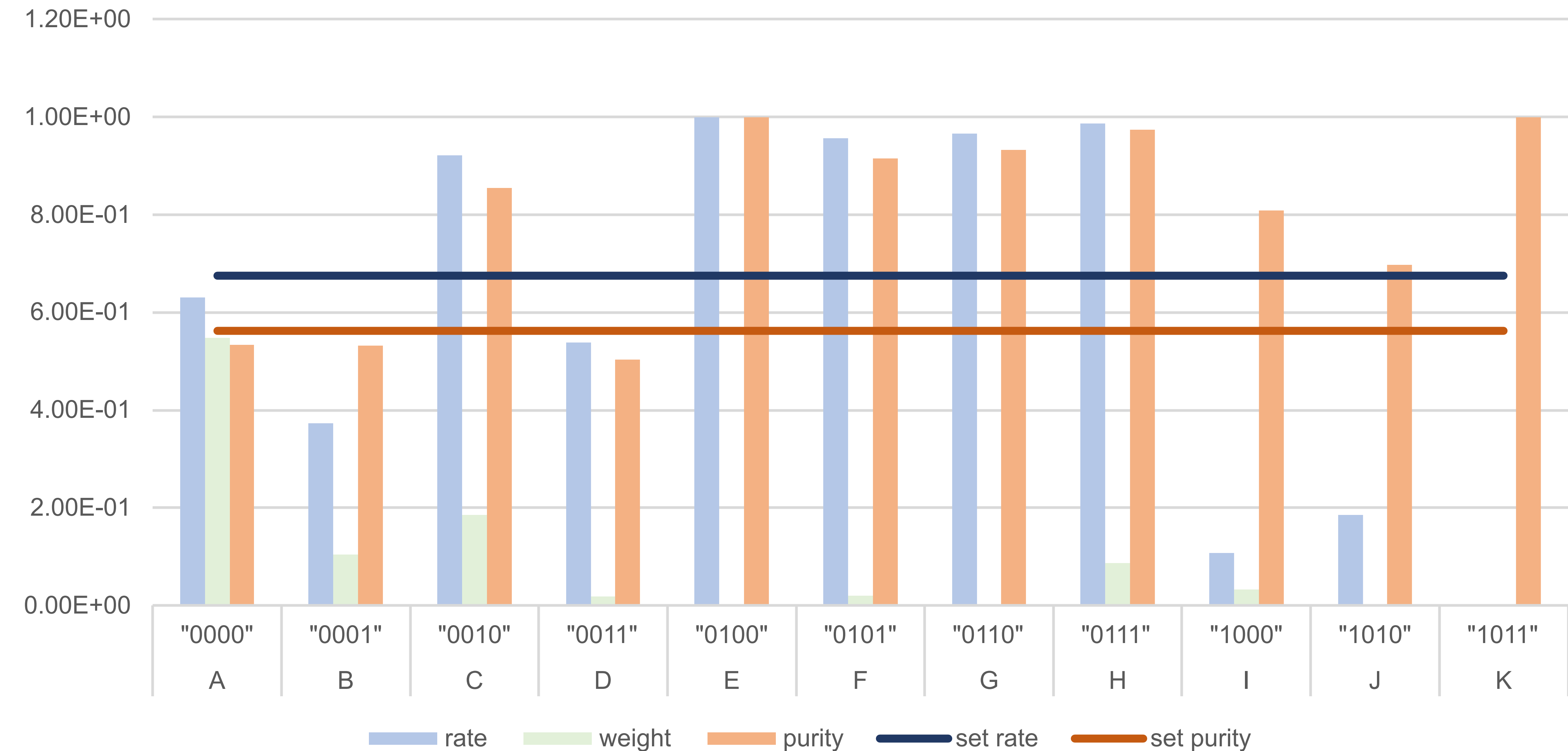
- | | | | |
|------------|--------------|-----------|-------------|
| Argentina | Dominican R. | Italy | Panama |
| Australia | Ecuador | Japan | Paraguay |
| Austria | Egypt | Kenya | Peru |
| Bangladesh | Estonia | Kuwait | Philippines |
| Belgium | Faroe I. | Lebanon | Poland |
| Bolivia | France | Lithuania | Reunion |
| Bosnia | French | Malawi | Romania |
| Brazil | Gabon | Malaysia | Russia |
| Bulgaria | Gambia | Maldives | Saudi A. |
| Cabo | Germany | Malta | Slovenia |
| Cambodia | Guatemala | Mauritius | South A. |
| Canada | Honduras | Mexico | Spain |
| Chile | Hong Kong | Nepal | Sri Lanka |
| China | India | New C. | Switzerland |
| Colombia | Indonesia | Nigeria | Thailand |
| Costa R. | Iran | Oman | Tunisia |
| Croatia | Iraq | Pakistan | Turkey |
| Cyprus | Ireland | Palestine | USA |
| Czech R. | Israel | | Zambia |

Supplementary Figure S4

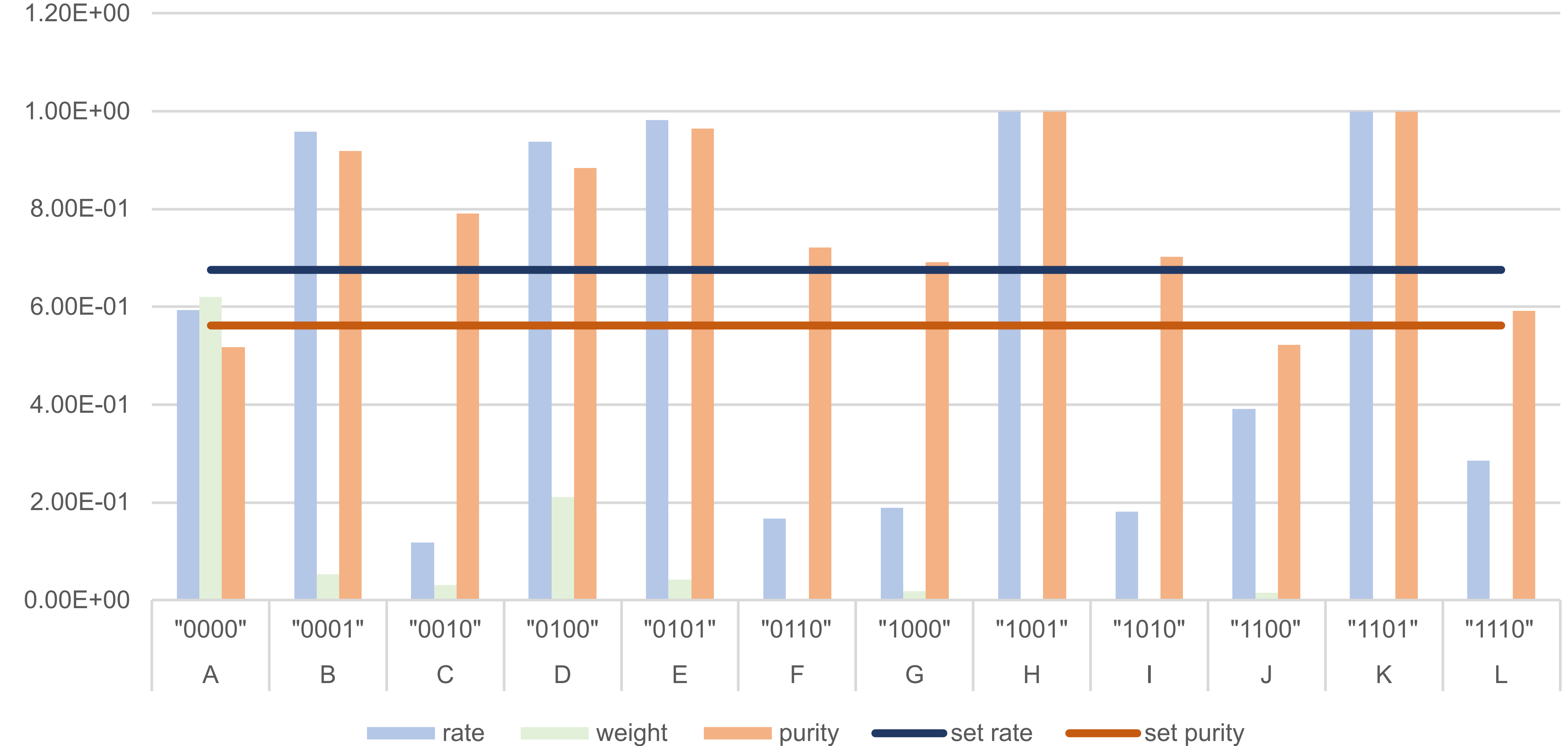


Supplementary Figure S5

4-SNV-1



4-SNV-2



Supplementary Figure S7

4-SNV-3

