THE UNIVERSITY OF ARIZONA
This research was supported by the NIH-NIGMS (R01GM127348 & R35GM149235 to RNG).

ConfuseNN: Interpreting convolutional neural network inferences in population genomics with data shuffling

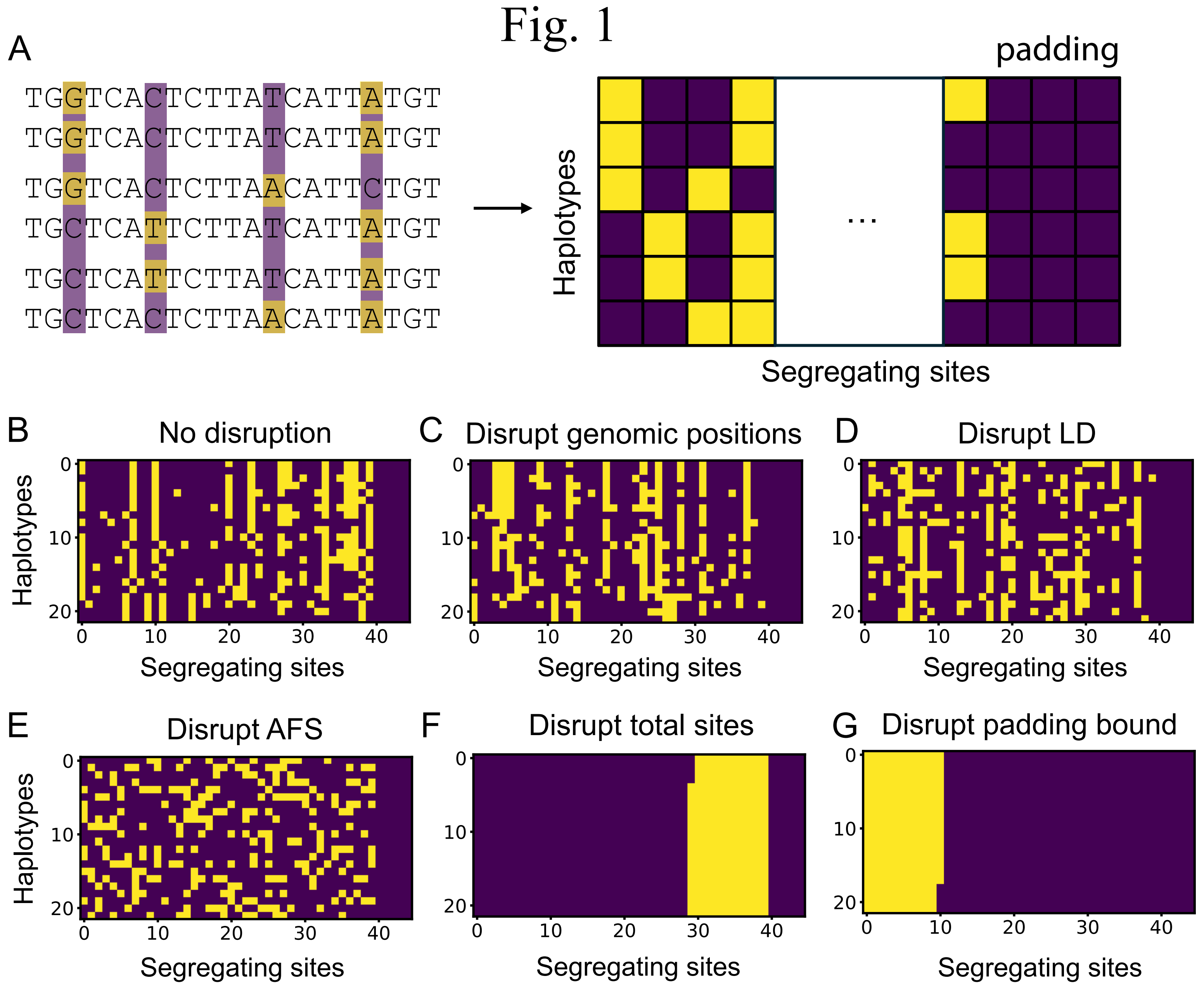
Linh N. Tran[1,2], David Castellano[2], Ryan N. Gutenkunst[2]
[1] Genetics Graduate Interdisciplinary Program, [2] Department of Molecular and Cellular Biology, University of Arizona; lnt@arizona.edu

Check out our preprint!

Significance

- Convolutional neural networks (CNNs) are powerful tools for population genomic inference, but understanding which genomic features drive performance remains challenging.
- We introduce ConfuseNN, a method that systematically shuffles input haplotype matrices to disrupt specific population genetic features and evaluate their contribution to CNN performance.

Fig. 1



Method

- The input data used by population genetic inference CNNs are 2D matrices with haplotypes as rows and segregating sites as columns, sometimes with padding (Fig. 1 A-B).
- By systematically shuffling this matrix, ConfuseNN sequentially removes signals of:
  - short- and long-range linkage disequilibrium (Fig. 1C-D),
  - allele frequency (Fig. 1E),
  - total diversity (Fig. 1F),
  - and padding bound (Fig. 1G).
- We use shuffled data to test three published CNNs trained for detecting positive selection (disc-pg-gan[1], Imagene[2]) and demographic history inference (Flagel et al.[3]).

Results

- Disc-pg-gan[1] is a generative adversarial network (GAN) whose discriminator is a CNN trained for binary classification (neutral vs. selection). When tested on shuffled data, this CNN performed similarly across tests, with the most drastic disruption resulting in only a slight performance reduction (Fig. 2A). This suggests that the original simulated test data may not have been challenging enough.

Fig. 2 Positive selection detection

disc-pg-gan

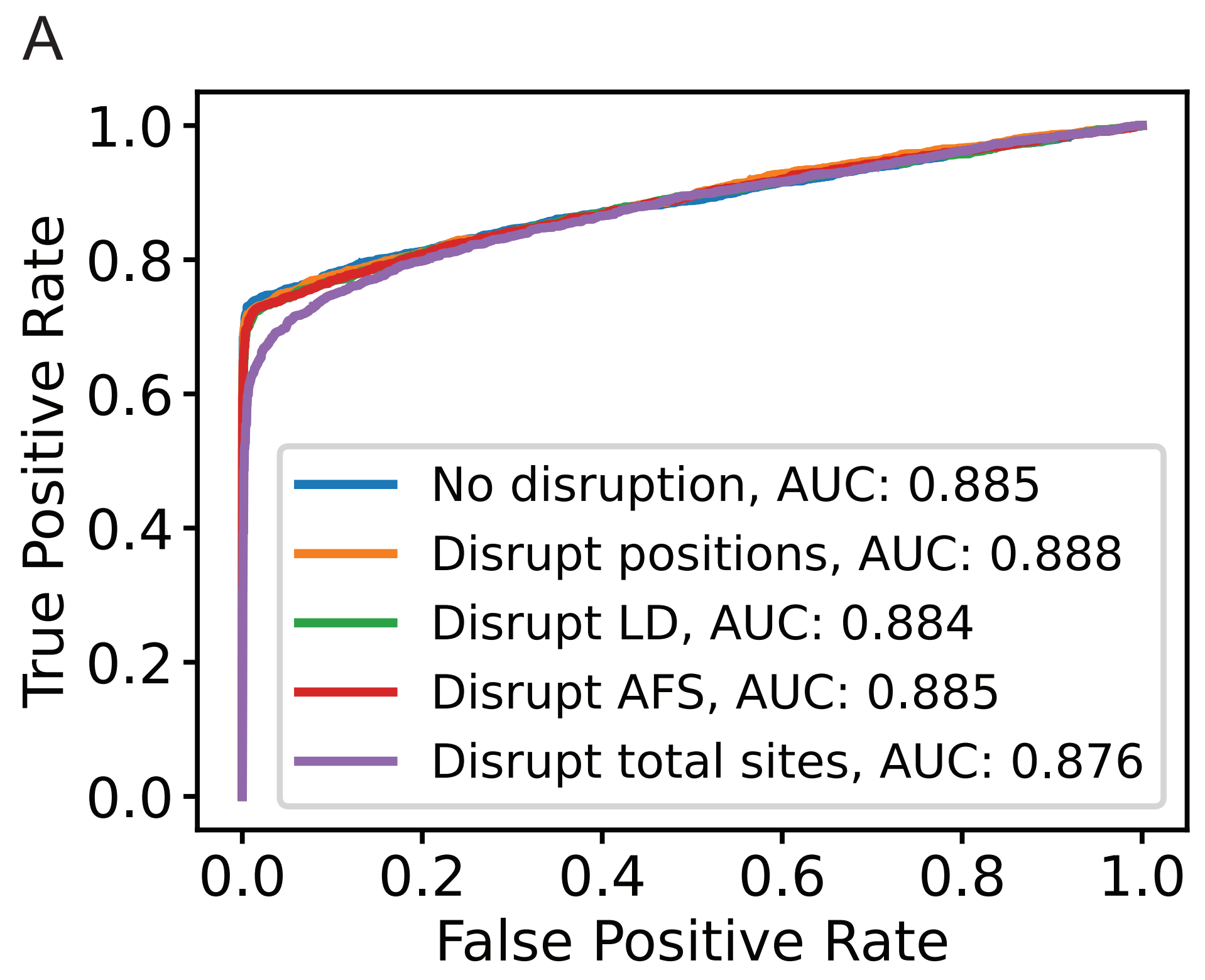
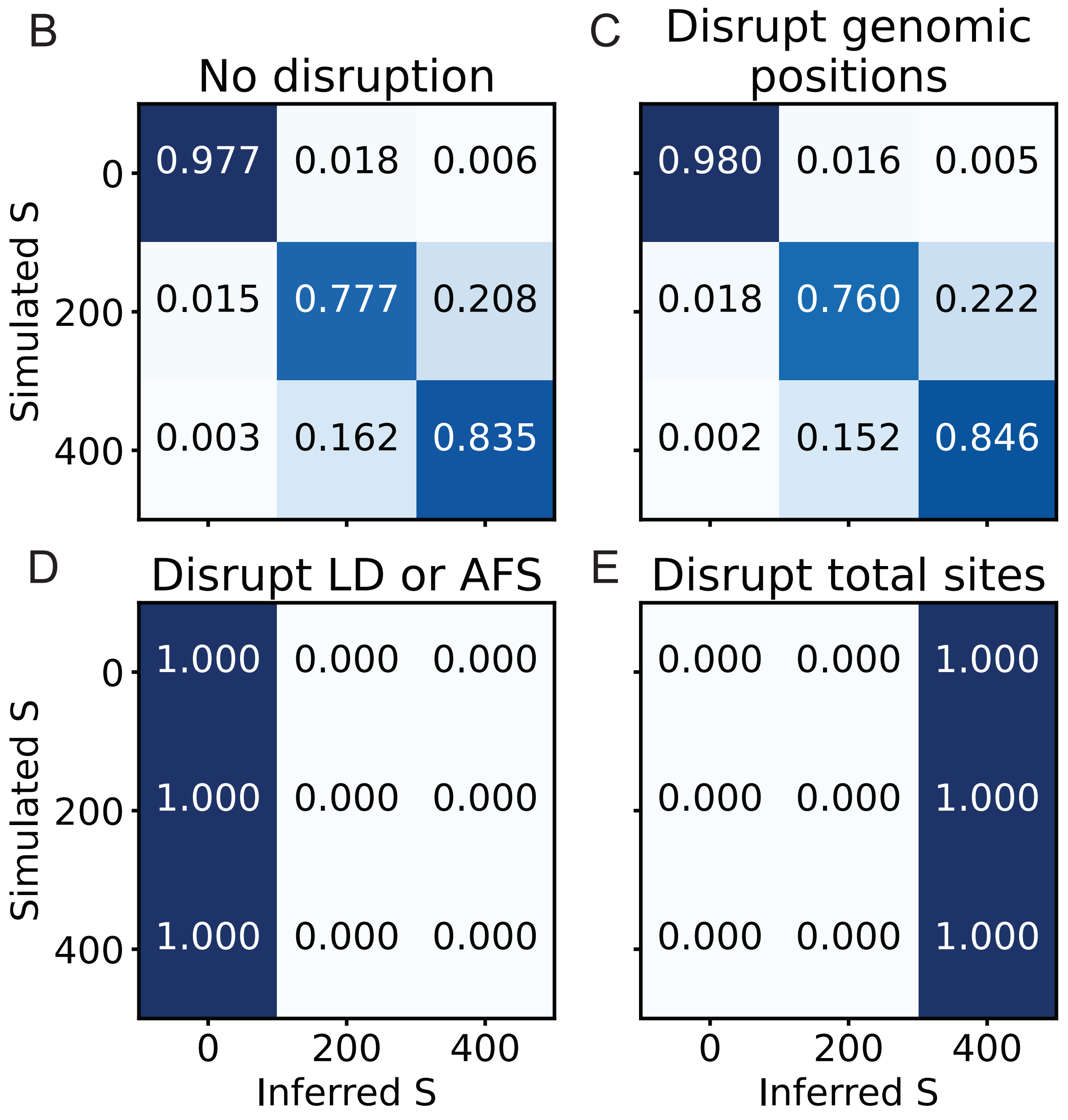
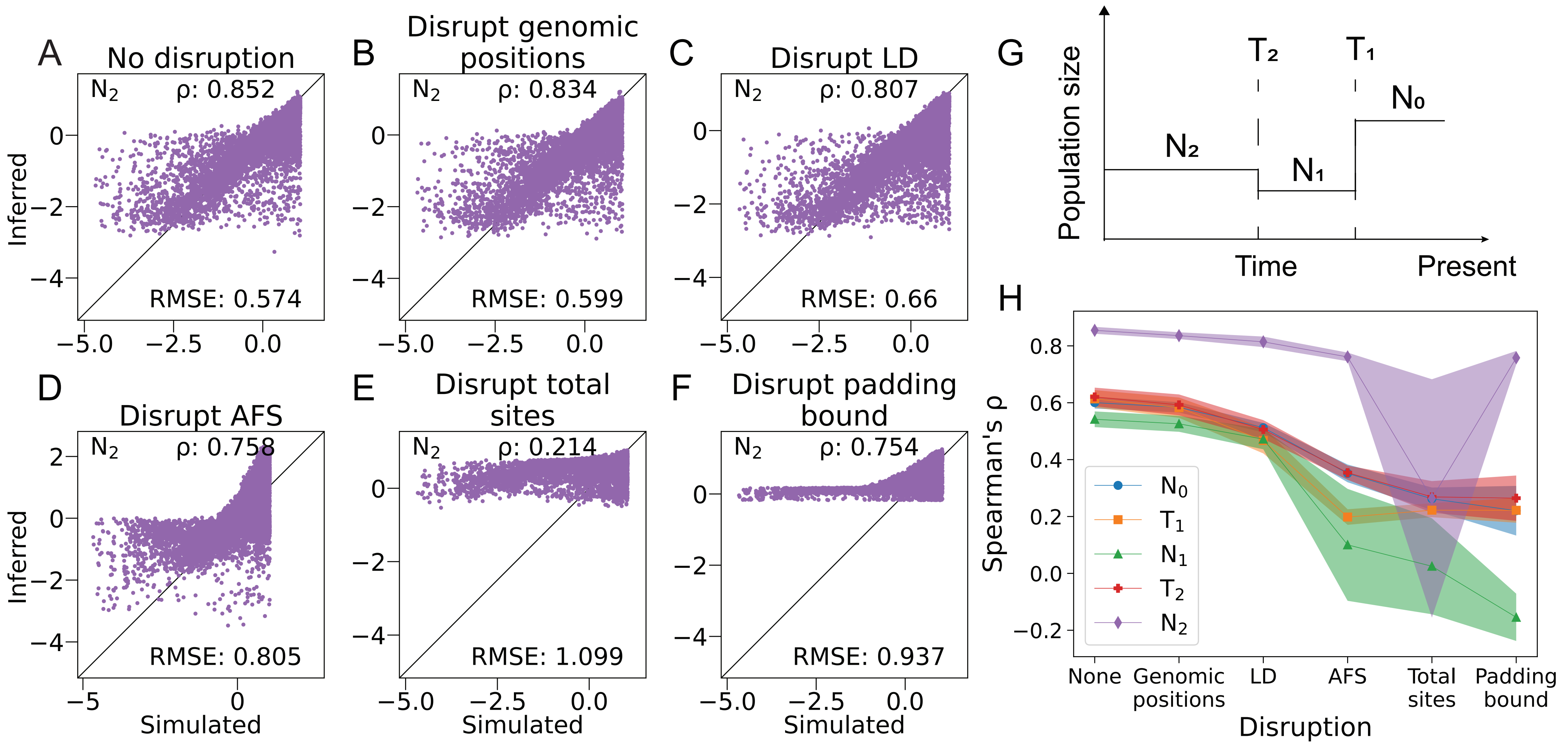


Fig. 2 Imagene



•Imagene[2] is a multiclass classification CNN (neutral S=0, moderate selection S=200, strong selection S=400). It breaks down when LD signals were disrupted, with further disruption at the allele frequency level did not change the pattern of failure (Fig. 2B-D). At the most drastic disruption level, the bias changed toward classifying all input data as strong selection instead of neutral (Fig. 2D-E). This is consistent with population genetic theory in that strong selection leads to a reduction in haplotype diversity, which the disrupted data at this level resembled. Imagene therefore likely relies on linkage features in the data to make its inference, consistent with the known importance of this signal for detecting positive selection.

Fig. 3 Demographic history inference



- Flagel et al.[3] developed a CNN to infer 5 parameters of a three-epoch demographic history model (Fig. 3G). Perturbation of linkage features did not greatly affect performance (Fig. 3A-C), but disrupting allele frequency, total diversity, and padding bound did (Fig. 3D-F). The pattern of performance degradation is consistent across ten independently trained instances of the CNN (Fig. 3H). This result is consistent with the CNN's convolutional kernel size being small (2x2), which was unlikely to strongly pick up linkage signals in the input data.

References

[1] Riley, R., Mathieson, I., & Mathieson, S. (2024). *Genetics*.
[2] Torada, L., et al. (2019). *BMC bioinformatics*.
[3] Flagel, L., Brandvain, Y., & Schrider, D. R. (2019). *MBE*.