MBE OXFORD

Efficient Detection and Characterization of Targets of Natural Selection Using Transfer Learning

Sandipan Paul Arnab [ID],[1] **Andre Luiz Campelo dos Santos** [ID],[1] **Matteo Fumagalli** [ID],[2,3]
Michael DeGiorgio [ID] [1,*]

[1]Department of Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL, USA
[2]School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK
[3]The Alan Turing Institute, London, UK

*Corresponding author: E-mail: mdegiorg@fau.edu.

Associate editor: Yuseob Kim

Abstract

Natural selection leaves detectable patterns of altered spatial diversity within genomes, and identifying affected regions is crucial for understanding species evolution. Recently, machine learning approaches applied to raw population genomic data have been developed to uncover these adaptive signatures. Convolutional neural networks (CNNs) are particularly effective for this task, as they handle large data arrays while maintaining element correlations. However, shallow CNNs may miss complex patterns due to their limited capacity, while deep CNNs can capture these patterns but require extensive data and computational power. Transfer learning addresses these challenges by utilizing a deep CNN pretrained on a large dataset as a feature extraction tool for downstream classification and evolutionary parameter prediction. This approach reduces extensive training data generation requirements and computational needs while maintaining high performance. In this study, we developed *TrIdent*, a tool that uses transfer learning to enhance detection of adaptive genomic regions from image representations of multilocus variation. We evaluated *TrIdent* across various genetic, demographic, and adaptive settings, in addition to unphased data and other confounding factors. *TrIdent* demonstrated improved detection of adaptive regions compared to recent methods using similar data representations. We further explored model interpretability through class activation maps and adapted *TrIdent* to infer selection parameters for identified adaptive candidates. Using whole-genome haplotype data from European and African populations, *TrIdent* effectively recapitulated known sweep candidates and identified novel cancer, and other disease-associated genes as potential sweeps.

Keywords: transfer learning, convolutional neural networks, logistic regression, population genomics, natural selection

Introduction

The study of biological diversity and adaptation relies on the identification and understanding of past evolutionary events. These events shed light on the processes that have shaped the genetics of organisms throughout history and those that are ongoing. Identifying genomic regions that are subject to the pressures of natural selection enables the discovery and isolation of genetic variants that may have contributed to adaptive traits in past environments. Understanding the genetic basis of vital characteristics like disease resistance, reproductive success, and physiological adaptations to changing environmental conditions illuminates the evolutionary history of a species. Predicting how species will respond to shifting weather patterns (Hoffmann and Sgrò 2011), altered habitats (Grant and Grant 2002), and the emergence of new diseases (Allison 1954) is enhanced by these findings. A window into the complex dynamics of evolution and adaptation can be found in the discovery of evolutionary events in genomes, which has implications for both theoretical and applied research.

A number of approaches have been taken to uncover traces of natural selection from polymorphisms within a population in light of neutral evolutionary processes, such as mutation,

recombination, migration, and genetic drift, which can both bolster and erode past adaptive signals (Kimura 1979; Slatkin 1987; Barton and Charlesworth 1998; Lynch 2010). Initial efforts for detecting such nonneutral regions involved summary statistics based on alterations of the distribution of allele frequencies, such as Tajima's D (Tajima 1989) and Fay and Wu's H (Fay and Wu 2003). These summary statistic approaches have been enhanced in contemporary studies to evaluate distortions in haplotype frequency distributions, such as H12 (Garud et al. 2015), and to assess the footprint of reduced genomic diversity through extended haplotype homozygosity methods, such as expected haplotype homozygosity (EHH) (Sabeti et al. 2002), integrated haplotype score (Voight et al. 2006), and nS_L (Ferrer-Admetlla et al. 2014). Further, more model-based approaches founded in population genetic theory or sensible probability distributions have revealed improved power, ability to isolate loci at which selection has likely acted, and effectiveness at estimating underlying parameters of the adaptive process (Nielsen et al. 2005; Stephens and Balding 2009; Chen et al. 2010; Pasaniuc et al. 2014; Vy and Kim 2015; Racimo 2016; Lee and Coop 2017; Lloyd-Jones et al. 2019; Harris and DeGiorgio 2020; Setter et al. 2020; DeGiorgio and Szpiech

2022). However, more recently, there has been a focus on a promising branch of selection detection methods based on machine learning or artificial intelligence that consider evidence from multiple input statistics or that operate on raw haplotype alignments or genotype calls (Schrider and Kern 2018; Korfmann et al. 2023).

Supervised machine learning represents an important branch of such methods, where models are trained on a set of observations often composed of a multidimensional input and a paired output value (Hastie et al. 2009). Training algorithms learn patterns and the functional relationship between the input and output data from these example observations (termed the training set) to hopefully make accurate output predictions on unseen future input data. Classification and regression tasks both make use of supervised learning algorithms to achieve their respective goals of organizing inputs into predefined categories (termed classes) and making continuous outcome predictions. There exist numerous supervised learning algorithms for classification and regression tasks, each with its own assumptions about data distributions and functional relationships (e.g. degree of nonlinearity) between the input and output values. Of course, no one approach is best, and the prediction abilities of the trained models highly depend on the correctness of the assumed data distribution and true underlying function relating the input and output. Artificial neural networks (ANNs) are universal approximators (Hornik 1991), having the capacity to represent any underlying function with enough model parameters, and can thus learn complicated patterns in a manner reminiscent of how the human brain works. One recent type of neural network that has enjoyed extensive success in image classification and recognition tasks is the convolutional neural network (CNN; Krizhevsky 2012), which is optimized for handling structured grid-like data.

Deployment of advanced machine learning and artificial intelligence tools has brought about a revolution in the realm of natural selection pattern recognition. Some of the applications of machine learning includes the extraction of features from one or many statistics computed in contiguous genomic windows, significantly increasing power to detect patterns compared to the classical usage of such statistics. Many of the state-of-the-art applications in this paradigm make use of powerful contemporary feature extraction tools on summary statistic arrays, ranging from simple linear to complex nonlinear models. Nevertheless, massive volumes of training data are typically necessary for nonlinear models, such as neural networks, to learn their parameters due to high model complexity. Such models are prone to overfitting and underperformance in the absence of large training sets, or through the use of automated frameworks to reduce the capacity or representation ability of the learned models (Hastie et al. 2009).

While data availability can be a limiting factor in some fields, population genetics benefits from the ability to generate effectively unlimited training data via simulation, particularly with simulation-on-the-fly techniques that provide new data with each iteration (Chan et al. 2018; Flagel et al. 2019; Torada et al. 2019; Battey et al. 2020), thus mitigating overfitting. However, minimizing the need for extensive training data remains beneficial to reduce computational demands and environmental impact, as highlighted by recent studies on the carbon footprint of bioinformatics tools (Grealey et al. 2022). Efficient and adaptable methods are crucial for analyzing genomic patterns shaped by natural selection. A framework that can swiftly summarize and capture these patterns without relying on application-specific summary statistics offers significant advantages. Utilizing a robust feature extractor capable of handling generalized genomic summaries ensures both adaptability and effectiveness, particularly in contexts with limited computational resources. This approach enhances the ability to detect and analyze complex genomic patterns, balancing the need for efficiency with sophisticated analysis.

This challenge can be addressed by applying transfer learning, a method of pretraining a neural network on unrelated data (Bozinovski 2020). Pretrained models are used in image classification to speed up convergence and improve outcomes (Hendrycks et al. 2019). The field of computer vision has been significantly enhanced by models trained on the ImageNet dataset (Deng et al. 2009). ImageNet is a massive collection of approximately 14 million annotated images organized into 1,000 categories that provide adequate diversity for developing powerful image categorization applications. Harnessing the diverse spectrum of images in this dataset and by leveraging high-performance computing resources, deep learning has made several prominent architectures the gold standard for image classification (Russakovsky et al. 2015; Beyer et al. 2020). A transfer learning model trained on a massive set of labeled images can be used as a feature extraction tool, thus eliminating the need to train complex models to succinctly represent input genomic data in a transformed space that is usable for downstream prediction tasks. Various transfer learning architectures have been successfully deployed in nonnatural image categorization tasks, such as medical imaging or industrial defect detection (Ming et al. 2021; Shafique et al. 2022; Suganyadevi et al. 2022), which contrasts with natural images that typically capture scenes from the natural world (i.e. landscapes or wildlife). This distinction underscores the robustness of these architectures, alleviating concerns about substandard performance when classifying images that capture patterns left by evolutionary events.

Instead of constructing a shallow CNN that operates on input images (Zhao et al. 2023) from a dataset with limited numbers of samples, it may be better to use pretrained deep CNN architectures. Unlike shallow CNNs, which are unable to capture the complex relationship between input images and the category they belong to, these pretrained deep CNNs are able to significantly improve the ability to model such complex relationships. Another advantage of pretrained CNNs is their capacity to generalize learning from one domain to another (Sharif Razavian et al. 2014; Yosinski et al. 2014). This ability helps pretrained CNNs perform better than shallow custom-made CNNs constructed and trained using smaller datasets (Azizpour et al. 2015; Tajbakhsh et al. 2016). Feature extraction from small datasets is enhanced by pretrained CNN architectures, leading to better generalization performance and resilience in tasks that have few data samples. Because training deep CNNs from scratch requires substantial computational resources, making use of pretrained models becomes a more feasible and efficient option.

In this article, we aim to develop powerful and robust natural selection detection and characterization tools by exploiting the feature extraction efficacy afforded by pretrained CNN models. Specifically, we apply five neural networks architectures pretrained on the large ImageNet database, which we use to extract transformed feature sets for input to classifiers to detect adaptive events and to regressors to predict

parameters of such events. By training models with a small number of simulated replicates and then testing their predictions with an array of confounding factors, we demonstrate that transfer learning can be a reliable tool for building predictive population genomic models that take images of haplotype variation as input without requiring extensive training data. Moreover, we introduce an efficient approach for constructing input images of haplotype variation that organizes diversity in a way that is easier for downstream image-based sweep classifiers to extract key patterns, ultimately improving their accuracies and powers. Here, we present *TrIdent* (*Tr*ansfer learning for *Ident*ification of adaptation), which is implemented as open-source software available at https://www.github.com/sandipanpaul06/TrIdent. As an empirical case study, we apply *TrIdent* to haplotype variation from a well-studied European human population (CEU; Utah residents with Northern and Western European ancestry) from the 1000 Genomes Project dataset, and are able to recapitulate established selective sweep candidates, such as *LCT* and *OCA2*, and discover candidate sweep signals in novel cancer-associated genes, such as *MTOR*, *LAMC2*, *RBMS3*, and *NKAIN2*, with high support. Similarly, when applied to a well-studied African human population (YRI; Yoruba in Ibadan) from the same dataset, *TrIdent* successfully identifies established sweep candidates, such as *HEMGN*, *SYT1*, *GRIK5*, *FOXP2*, and *APOL1*, while also uncovering novel candidate sweep signals in *ROBO2*, with strong support.

Results

Modeling Description

We used the coalescent simulator discoal (Kern and Schrider 2016) under two nonequilibrium demographic histories estimated for European (CEU) and sub-Saharan African (YRI) humans (Tennessen et al. 2012), which respectively include a recent severe population bottleneck and a recent population expansion, to simulate neutral and sweep replicates to train and evaluate *TrIdent*. We simulated sweeps with per-generation selection strengths (s) ranging from 0.005 to 0.1, beneficial allele frequencies (f) from $1/(2N_e)$ to 0.2, and fixation times of beneficial alleles (τ) from 0 to 2,000 generations prior to sampling. Moreover, sweep and neutral replicates were generated with per-site per-generation mutation rates (μ) from 2.21×10^{-9} to 2.21×10^{-8}, and per-site per-generation recombination rates (r) from 0 to 3×10^{-8} with a mean rate of 10^{-8} (see *Simulation protocol* subsection of the *Materials and Methods*). Together, these settings and parameters allowed for the examination of a broad range of genetic and adaptive conditions, while focusing on two distinct demographic histories. Given the wide range of selection strengths and beneficial allele frequencies, along with varying mutation and recombination rates and fixation events that may have occurred far in the past, and further complicated by fluctuations in population size, substantial overlap in the distribution of genetic variation between neutral and selection settings is expected. This overlap creates significant challenges in distinguishing the sometimes subtle genetic patterns driven by positive selection from those that emerge under neutrality.

For the purpose of producing inputs to feed *TrIdent*, we formed the CEU and YRI training datasets by using 1,000 neutral and 1,000 sweep replicates that were simulated using their respective demographic histories. In addition, we generated two sets of test and validation datasets to complement the CEU and YRI training datasets, each consisting of 1,000 neutral and 1,000 sweep replicates. We detail the processing of input images from these simulated replicates in the *Image Generation* subsection of the *Methods*. The image generation process (Fig. 1; top panel) transforms these replicates into grayscale images representing sorted minor allele counts across shifting small and overlapping windows, and then resized to match the input expectations of the pretrained models.

Heatmaps depicting mean images for the sweep and neutral classes for the CEU and YRI datasets show that, in contrast to the neutral image, the sweep image features a dark vertical segment in the central columns representing a loss of diversity in the central columns due to high-frequency haplotypes (supplementary fig. S1, Supplementary Material online). That is, these regions contain a string of major alleles that are at high frequency near the center of the sweep replicates. Once these input images have been resized, each pixel is standardized according to its value at that particular position across all training images, so that each pixel has a mean of zero and a standard deviation of one. Heatmaps depicting mean standardized neutral and sweep images show that the neutral image has a red segment of positive values in the central columns to complement the blue segment of negative values in the sweep image (Fig. 2). Thus, standardization reveals a clear distinction between the two classes, and serves as a proof of concept that the applied image generation technique presents a pattern that can be employed to discriminate between positive selection and neutrality.

These standardized images are fed to pretrained models, after which we retain the output vector, representing transformed features or embeddings, from the global average pooling (GAP) layer that we attach to the pretrained model (Fig. 1; bottom panel). The GAP layer computes the mean value of each feature map, effectively reducing the spatial dimensions and summarizing the most salient features of the image. By averaging each feature map into a single node, GAP reduces the number of parameters in the model and makes the model more robust to spatial translations, offering advantages over traditional fully connected layers (Szegedy et al. 2015). A penalized logistic regression classifier is then trained using the GAP layer outputs from 1,000 neutral and 1,000 sweep simulations, along with an elastic-net regularization penalty to simultaneously modulate model complexity and sparsity (Zou and Hastie 2005; Hastie et al. 2009). In particular, the magnitudes of model parameters associated with each transformed input feature are controlled by an L_2-norm penalty, and feature selection is carried out to control sparsity by setting some model parameters to zero through an L_1-norm penalty. The hyperparameter $\alpha \in \{0.0, 0.1, \ldots, 1.0\}$ specifies the proportion of model regularization that derives from the L_2-norm penalty, whereas the hyperparameter $\lambda \in \{10^{-6}, 10^{-5}, \ldots, 10^5\}$ designates the amount of total regularization. We utilized the LogisticRegression module from the *Scikit-learn* (Pedregosa et al. 2011) package in Python to train the classifier. The best pair of hyperparameters to train the classifier was determined based on binary cross-entropy loss across the validation set.

Choosing an Appropriate Pretrained Model

Numerous pretrained models are available for transfer learning, including *VGGNet*s, *GoogleNet*s, *ResNet*s, and *EfficientNet*s that each have their own architectures and distinguishing features. A minimalist architecture and a modest
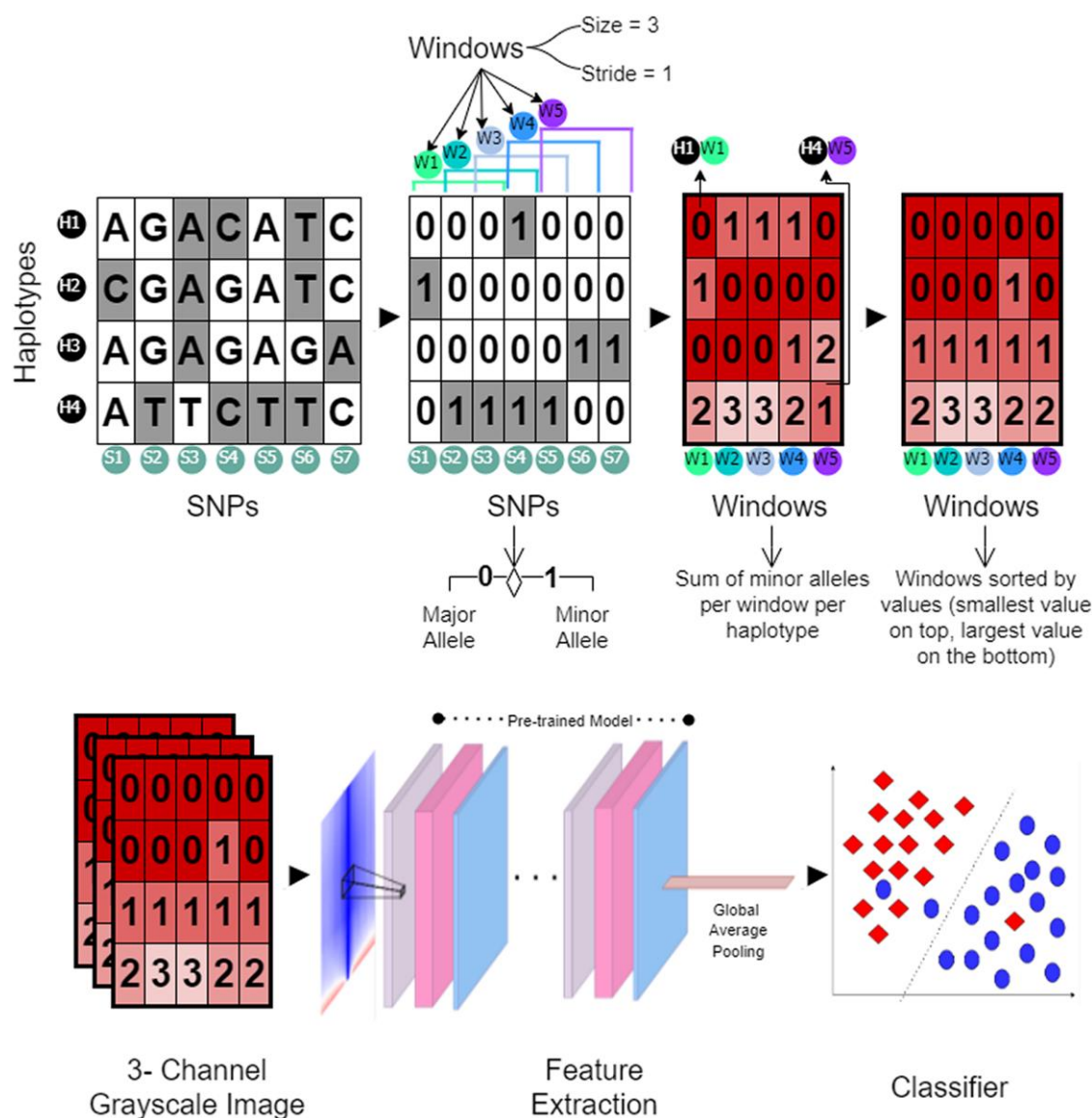
Fig. 1. Depiction of the *TrIdent*[*IRV2*] model (bottom panel), including the native *TrIdent* image generation method (top panel). As described in *Image generation* subsection of the *Methods*, for a haplotype alignment with haplotypes on rows and SNPs on columns, the major allele is represented by zero and the minor allele by one at each SNP. From this processed alignment, the number of minor alleles for each haplotype is counted in a window and window locations are shifted by a specific stride, which is chosen as window size of three SNPs and a stride of one SNP in this schematic. The minor allele counts for each window are then sorted, such that the top row with have smallest value and the bottom row the largest value for a given window. A matrix based on a certain number of consecutive windows (here five windows) is created, this matrix is copied over two more channels to create a tensor, resulting in a three-channel grayscale image. This image is fed as input to the "Feature Extraction" block consisting of a pretrained deep CNN model that may incorporate various combinations (indicated by blocks of different colors) of a subset of the following layers: convolutional, maxpooling, dense, dropout, squeeze-and-excitation, depth-wise separable convolutional, and residual connections. A GAP layer is attached to the pretrained model to generate a feature vector, which is then used to train a classifier. The *TrIdent*[*IRV2*] model, which is focused on this article, combines the use of *InceptionResNetV2* as the pretrained model and penalized logistic regression as the binary classifier.

number of convolutional filters distinguish the Visual Geometry Group (*VGG*)*Net* from other models pretrained on the ImageNet dataset (Simonyan and Zisserman 2015). Szegedy et al. (2015) found that the Inception modules of *GoogleNet* improved performance and allowed it to capture features at multiple scales, whereas *ResNet* mitigated the vanishing gradient problem, a hurdle in training extremely deep networks consisting of multiple layers designed to learn complex hierarchical representations of the input data by focusing on residual learning, which helps information flow more smoothly through the layers via skipped connections between layers (He et al. 2016). Recent developments like *EfficientNet* and *MobileNet* employ a compound scaling parameter,

through which the numbers of layers and nodes in a layer are controlled. Capacity to modulate the depths of layers and width of nodes allows these models to offer a tradeoff between efficiency and complexity (Tan and Le 2019). The *InceptionResNet* architecture is an augmentation of the *Inception* architecture by incorporation of residual learning (Szegedy et al. 2017), and represents a robust image recognition system with high accuracy in numerous tasks. By adding residual connections to its Inception blocks, *Inception ResNetV2* improves network information flow and gradient propagation during training (Wang et al. 2021). We first study the efficacy of five pretrained models (*InceptionResNetV2*, *VGG16*, *EfficientNetB0*, *EfficientNetB7*, and *MobileNetV2*)
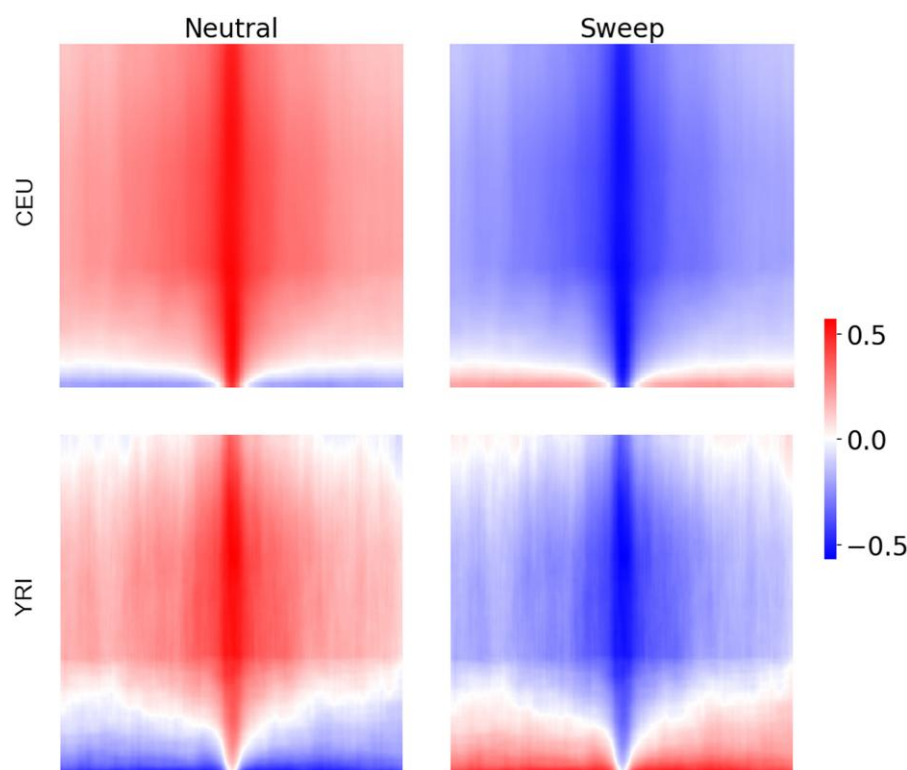
Fig. 2. Heatmaps depicting standardized input images of size 224 × 224, averaged across the 1,000 training replicates for either the neutral or sweep class simulated under either the European (CEU) or Sub-Saharan African (YRI) human demographic history (Tennessen et al. 2012). Standardized input images are processed as in the *Image generation* subsection of the *Materials and Methods*, with standardization occurring across the 2,000 neutral and sweep training replicates for each pixel. Rows of the images represent haplotypes, whereas columns represent genomic window of 25 contiguous SNPs within a haplotype, with an equal number of windows flanking the center of a simulated genomic region. The colorbar indicates a measure proportional to the number of minor alleles within the haplotype window relative to the mean number (scaled by the standard deviation of minor allele counts) for that window across the neutral and sweep training observations. Darker blue shading represents a higher number of major alleles than average and darker red shading represents a higher number of minor alleles than average.

as complex feature extraction methods to develop adaptive event classification tools. We accessed these models with TensorFlow (Abadi et al. 2015), as their APIs make communication with these pretrained models as well as their integration into workflows easier.

We evaluated the accuracy and power of the *TrIdent* logistic regression classifiers developed based on GAP layer outputs attached to the pretrained *InceptionResNetV2*, *VGG16*, *EfficientNetB0*, *EfficientNetB7*, and *MobileNetV2* on both the CEU and YRI datasets. For convenience of presentation, we respectively refer to these models as *TrIdent*[*IRV2*], *TrIdent*[*VGG16*], *TrIdent*[*ENB0*], *TrIdent* [*ENB7*], and *TrIdent*[*MNV2*]. On the CEU dataset, we found that *TrIdent*[*IRV2*], with its high number of parameters and computational demand, achieves the highest classification accuracy of 86.2% along with the highest sweep detection accuracy among all five classifiers (supplementary fig. S2, Supplementary Material online). In contrast, the most lightweight and computationally efficient *TrIdent*[*VGG16*] as well as the moderately accurate and moderately efficient architecture of *MobileNetV2* achieve identical accuracies of 85.1% (supplementary fig. S2, Supplementary Material online). The two *EfficientNet* classifiers, *TrIdent*[*ENB0*] and *TrIdent*[*ENB7*], are the worst performers out of the five *TrIdent* models. However, *TrIdent*[*ENB0*], with an accuracy of 84.95%, outperforms *ENB7*, which achieves an accuracy of 83.8%. This finding suggests that the shallower architecture of *ENB0* performs better in this case, contrary to the expectation

that deeper architectures typically yield higher accuracy. These results are mirrored by the powers of these methods to detect sweeps based on receiver operating characteristic (ROC) curves, with *TrIdent*[*IRV2*], *TrIdent*[*MNV2*], and *TrIdent*[*VGG16*] exhibiting similar ROC curves and *TrIdent*[*IRV2*] slightly edging out *TrIdent*[*MNV2*] and *TrIdent*[*VGG16*] (supplementary fig. S2, Supplementary Material online). On the other hand, *TrIdent*[*ENB7*] showcases the lowest power to detect sweeps among the five classifiers (supplementary fig. S2, Supplementary Material online).

Our results also show that the classification behavior of *TrIdent* on the YRI dataset relative to CEU depends on network architecture (compare supplementary figs. S2 and S3, Supplementary Material online). Specifically, *TrIdent*[*IRV2*] and *TrIdent*[*MNV2*] lead the set of classifiers with overall accuracies of 87.55% and 87.50%, respectively, with *TrIdent*[*IRV2*] marginally edging out *TrIdent*[*MNV2*] (supplementary fig. S3, Supplementary Material online), similarly to the application on the CEU dataset (supplementary fig. S2, Supplementary Material online), with slightly higher accuracies than on the CEU dataset, which may be expected as the YRI dataset has higher mean neutral haplotype diversity. The ROC curves of this pair of classifiers also indicate their proximity in performance metrics, while clearly distancing themselves from the other three classifiers. Moreover, *TrIdent*[*ENB0*] and *TrIdent*[*ENB7*] again prove to be the weakest of the set of classifiers with accuracies of 74.35% and 77.30%, respectively (supplementary fig. S3,

Supplementary Material online). However, unexpectedly, their accuracies are significantly lower than on the CEU dataset. *TrIdent*[*VGG16*], similar to the CEU test case, stations itself in the middle of these pairs of classifiers with an overall accuracy of 84.55% (supplementary fig. S3, Supplementary Material online). Following the superior classification performance of *TrIdent*[*IRV2*] compared to the other *TrIdent* models in both the CEU and YRI test cases, we elected to pursue *TrIdent*[*IRV2*] for future analyses. Given our simulation setup and the two classes analyzed, *TrIdent*[*IRV2*] proved to be the most effective and consistent model. However, it may not be the best choice for all applications. Other pretrained models, including but not limited to the ones we tested, could perform better depending on the specific context and requirements.

Viability of Alternate Architectures and Methods

Though *TrIdent*[*IRV2*] uses a linear model (logistic regression classifier) with features extracted from the *InceptionResNetV2* architecture, we also evaluated nonlinear models based on ANNs for classification using the same extracted features. We refer to this approach as *TrIdent*[*IRV2, ANN*]. We optimized the *TrIdent*[*IRV2, ANN*] architecture by varying the number of hidden layers $L \in \{1, 2, 3, 4, 5\}$, the number of nodes $n_\ell \in \{100, 200, \ldots, 1{,}000\}$ in each hidden layer $\ell \in \{1, 2, \ldots, L\}$, and the activation function for nodes in each hidden layer $\phi_\ell \in \{$ReLU, sigmoid, tanh$\}$. The node in the output layer has a sigmoid activation function, and the input layer has 1,536 nodes, which is identical to the number of features generated by the GAP layer attached to the *InceptionResNetV2* architecture. Like our linear *TrIdent*[*IRV2*] model, *TrIdent*[*IRV2, ANN*] also employs L_1—and L_2-norm regularization penalties for each hidden and output layer, with regularization hyperparameters chosen from the same ranges as those for the linear model (see *Modeling description* subsection). Each model is trained to minimize binary cross-entropy loss using the Adam optimizer (Kingma and Ba 2015). The optimal architecture and hyperparameters were selected based on the smallest binary cross-entropy validation loss on the CEU dataset, which resulted in a model of $L = 3$ hidden layers with $n_1 = 1{,}000$, $n_2 = 500$, and $n_3 = 500$ nodes in the first, second, and third hidden layers, respectively, sigmoid activations for nodes in the first and second hidden layers, and a tanh activation for nodes in the third hidden layer. We obtained a near identical optimal architecture for the YRI dataset, with the exception of an additional hidden layer ($L = 4$) with $n_4 = 500$ nodes and a tanh activation. However, when applied to the same test data used for the linear model, *TrIdent*[*IRV2, ANN*] classification performance was virtually unchanged relative to *TrIdent*[*IRV2*] (supplementary fig. S4, Supplementary Material online). This result was consistent even with various training set sizes (1,000, 3,000, and 5,000 observations per class), suggesting that the nonlinearity introduced by the ANN did not lend additional benefit beyond that of the *InceptionResNetV2* feature extractor. We therefore decided to continue with the linear *TrIdent* model for classification tasks, as the simpler logistic regression model has fewer parameters and is faster to train than the ANN.

We also tested the classification performance of *TrIdent*[*IRV2*] against *T-REx* (Amin et al. 2023), which is a high-performing tensor decomposition-based sweep classifier that operates on a windowed aggregation of genetic diversity as its input, similar to our image generation method, and that uses an L_1—and L_2-norm penalized logistic regression classifier for prediction, which is identical to *TrIdent*. To provide a fair comparison of performance, we tested *TrIdent*[*IRV2*] and *T-REx* with both image generation techniques. That is, *TrIdent*[*IRV2*] and *T-REx* were trained and evaluated based on their own native image generation approaches, as well as the image generation approaches of the competing model. For clarity, we denote the *TrIdent*[*IRV2*] model trained and tested on the alternate image generation style of *T-REx* as *TrIdent*[*IRV2, alt*], and the *T-REx* model trained and tested on the alternate image generation style of *TrIdent* as *T-REx*[*alt*]. We find that on the CEU dataset, the *TrIdent*[*IRV2, alt*] alternate implementation loses 2% accuracy compared to the stock *TrIdent*[*IRV2*] model (Fig. 3). Moreover, *T-REx* trained on its native image generation style exhibits lower classification accuracy (79.6%) compared to both the *TrIdent*[*IRV2*] models (Fig. 3). However, the *T-REx*[*alt*] alternate implementation reports a higher classification accuracy (81.75%) compared to the stock *T-REx* model, but falls short of *TrIdent*[*IRV2*] by 4.45% (Fig. 3). The powers of each of these models to detect sweeps at a 5% false positive rate (FPR), lend further support that *TrIdent* models significantly outperform the *T-REx* models, regardless of the input image representation used by *T-REx* (Fig. 3). Specifically, *TrIdent*[*IRV2*] marginally surpasses *TrIdent*[*IRV2, alt*] as the most powerful model of the four at a 5% FPR.

A parallel illustration is observed when the performances of the stock and alternate models of *TrIdent*[*IRV2*] and *T-REx* are evaluated on the YRI dataset (Fig. 4). The stock *TrIdent*[*IRV2*] model leads the four models with an accuracy of 87.55%, followed by *TrIdent*[*IRV2, alt*] (85.30%), *T-REx*[*alt*] (83.85%), and *T-REx* (82.80%). The detection capability of *TrIdent*[*IRV2*] for identifying sweeps at a 5% FPR provides additional evidence that *TrIdent*[*IRV2*] is the superior model of the four in this test case (Fig. 4). Though a deeper inspection of model performances shows that the performance difference seen among the models in the CEU test case (Fig. 3) is noticeably reduced in the YRI test case (Fig. 4), with the absence of a population bottleneck in the YRI dataset apparently bridging the performance gap among models.

In addition to comparing *TrIdent*[*IRV2*] with *T-REx*, we conducted a comprehensive evaluation by also comparing it against diploS/HIC (Kern and Schrider 2018), a state-of-the-art selective sweep detector that integrates summary statistics with comparatively shallow CNNs. We trained and tested diploS/HIC in its native setting, with one modification as we calculated the summary statistics in 101 windows instead of 11. While diploS/HIC typically uses 11 windows (resulting in 11 features for the downstream CNN model for each summary statistic), we chose to use 101 windows (resulting in 101 features for the downstream CNN model for each summary statistic), which retains the focal window but within an increased number of flanking windows. This adjustment resulted in a finer representation, and enhanced the prediction performance of diploS/HIC in our initial investigation. We also extended our analysis to evaluate a variant of the *TrIdent*[*IRV2*] architecture to better understand its performance when operating on summary statistic data, which we denote as *TrIdent*[*IRV2, SS*]. Specifically, we employed the 12 diploS/HIC summary statistics—$\hat{\pi}$ (Tajima 1983), $\widehat{\theta}_W$ (Watterson 1975), Tajima's D (Tajima 1989), the variance, the skewness, and the kurtosis of multilocus genotype distances between pairs of sampled individuals, the multilocus
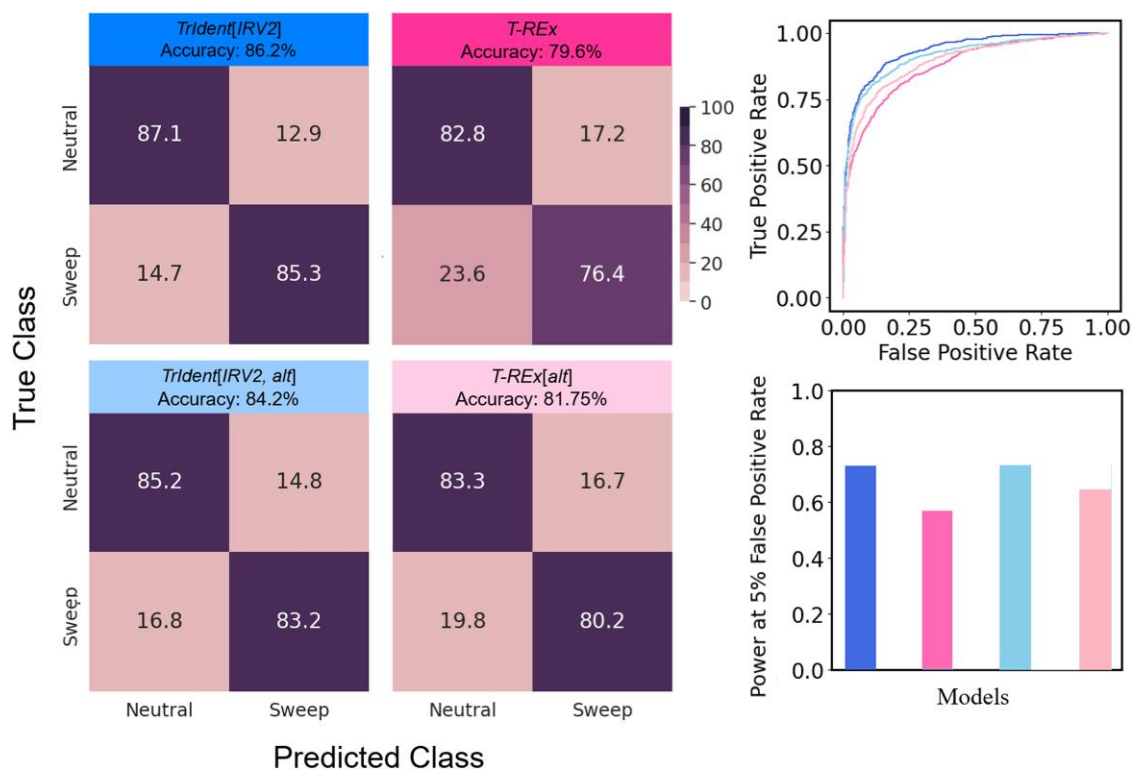
Fig. 3. Classification rates and accuracies as depicted by confusion matrices, powers (true positive rates) to detect sweeps as depicted by ROC curves to differentiate sweeps from neutrality, and powers at a 5% FPR to detect sweeps on the CEU dataset for the best performing *Trident* model (*TrIdent[IRV2]*) compared to *T-REx*. The comparison also includes *TrIdent[IRV2, alt]* and *T-REx[alt]*, which are trained and tested using their alternate image generation styles (see *Viability of alternate architectures and methods* subsection of the *Results* for details).

Fig. 4. Classification rates and accuracies as depicted by confusion matrices, powers (true positive rates) to detect sweeps as depicted by ROC curves to differentiate sweeps from neutrality, and powers at a 5% FPR to detect sweeps on the YRI dataset for the best performing *Trident* model (*TrIdent[IRV2]*) compared to *T-REx*. The comparison also includes *TrIdent[IRV2, alt]* and *T-REx[alt]*, which are trained and tested using their alternate image generation styles (see *Viability of alternate architectures and methods* subsection of the *Results* for details).
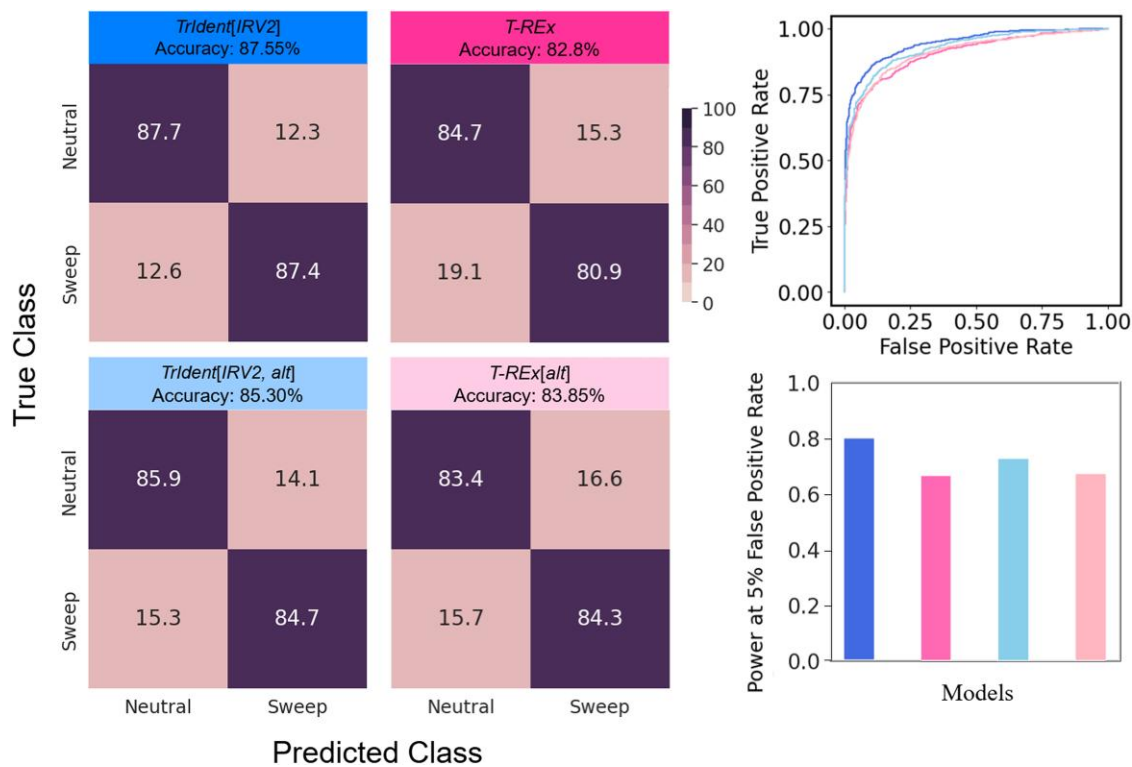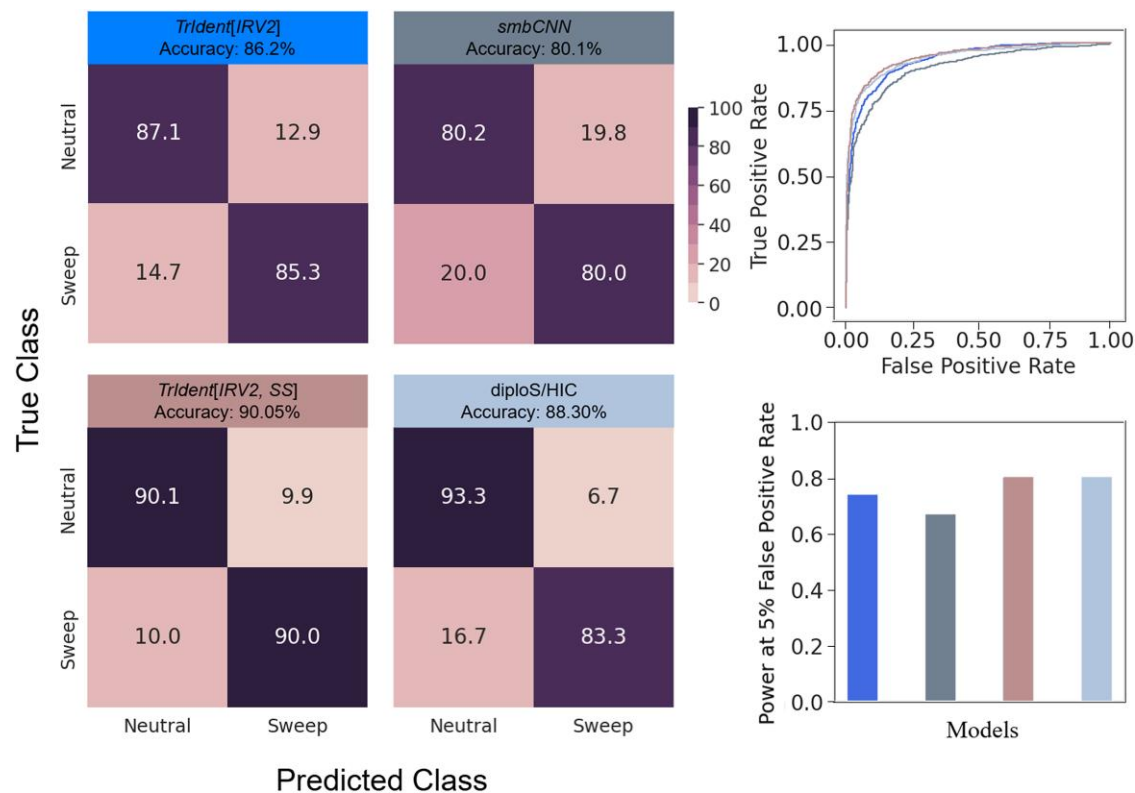
Fig. 5. Classification rates and accuracies as depicted by confusion matrices, powers (true positive rates) to detect sweeps as depicted by ROC curves to differentiate sweeps from neutrality, and powers at a 5% FPR to detect sweeps on the CEU dataset for the best performing *TrIdent* model (*TrIdent*[*IRV2*]) compared to diploS/HIC and *smbCNN*. The *smbCNN* model represents a custom-built shallow CNN trained using *TrIdent*'s native images, and *TrIdent*[*IRV2, SS*] represents an alternate *TrIdent*[*IRV2*] architecture trained using summary statistics based images (see *Viability of alternate architectures and methods* for details).

genotype equivalents of EHH H_1, H_{12} (Garud et al. 2015; Harris et al. 2018), and H_2/H_1 (Garud et al. 2015; Harris et al. 2018), unphased Z_{ns} (Kelly 1997; Rogers and Huff 2009), and the maximum value of ω (Kim and Nielsen 2004; Rogers and Huff 2009)—which provide insights into multiple aspects of genetic diversity affected by selective sweeps, including the structure of haplotype variation, nucleotide diversity, and linkage disequilibrium. Together, these components offer a comprehensive view of the effects of selective sweeps on genomic regions (Schrider and Kern 2016; Kern and Schrider 2018). We computed these summary statistics in 101 contiguous and nonoverlapping windows (each of length 11 kb) across simulated sequences of length 1.1 Mb, resulting in input matrices of dimension 12×101, with summary statistics on the rows and computed values across windows along the columns. We then resized these matrices to dimension 299×299 to match the expected input image size of *InceptionResNetV2*. Moreover, though diploS/HIC was originally designed to differentiate among five classes, we retooled it to differentiate between two classes, as in other studies (Mughal et al. 2020; Arnab et al. 2023).

Next, we evaluated an alternate version of diploS/HIC by training a shallow CNN using the images produced by the native image generation method of *TrIdent*, which we refer to as *smbCNN* (shallow multibranch CNN). We employed the identical CNN architecture used in diploS/HIC (Kern and Schrider 2018) to extract features from input images. This shallow architecture is comprised of three branches, each containing two convolution layers with ReLU activation (Nair and Hinton 2010) as well as max pooling (LeCun et al.

1998), dropout (Srivastava et al. 2014), and flatten layers, with the kernel sizes and dilation rates (Yu and Koltun 2015) utilized in the convolution layers determining the differences among these three branches. These branches are then concatenated prior to adding two pairs of dropout and dense layers to form the final model.

By examining *TrIdent*[*IRV2, SS*] and *smbCNN*, we aim to understand the potential impact of incorporating summary statistics into the *TrIdent*[*IRV2*] architecture, as well as the effect of using shallow CNNs trained on *TrIdent* native images. Comparing *TrIdent*[*IRV2*], *TrIdent*[*IRV2, SS*], diploS/HIC, and *smbCNN* across the CEU (Fig. 5) and YRI (Fig. 6) datasets, we can assess the impact of different architectural choices on model performance. In both demographic test cases, *smbCNN* was the poorest performing model among the four, with accuracies of 80.1% and 83.7% on the CEU and YRI datasets, respectively (Figs. 5 and 6). However, the two models employing summary statistics, *TrIdent*[*IRV2, SS*] (accuracy of 90.05% on CEU and 94.95% on YRI) and diploS/HIC (accuracy of 88.30% on CEU and 94.15% on YRI), consistently demonstrate higher classification accuracy compared to *TrIdent*[*IRV2*] (Figs. 5 and 6). *TrIdent*[*IRV2, SS*] outperformed diploS/HIC in both cases. Prior to resizing, summary statistic-based images are significantly smaller than native *TrIdent* images (12×101 compared to 198×499 for CEU or 216×499 for YRI). Because different summary statistics computed within a given window are arranged along a column in consecutive rows, some correlation among values is anticipated. Moreover, values along rows, which are individual summary statistics computed across consecutive windows,
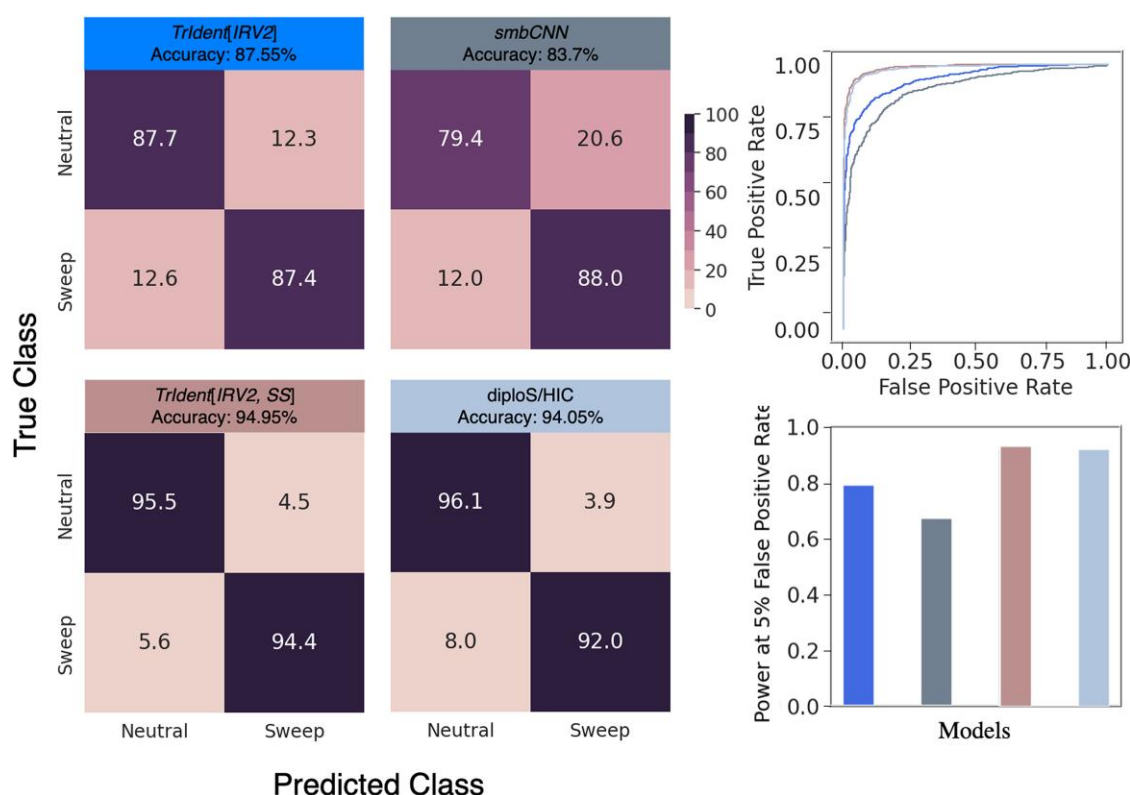
Fig. 6. Classification rates and accuracies as depicted by confusion matrices, powers (true positive rates) to detect sweeps as depicted by ROC curves to differentiate sweeps from neutrality, and powers at a 5% FPR to detect sweeps on the YRI dataset for the best performing *TrIdent* model (*TrIdent*[*IRV2*]) compared to diploS/HIC and *smbCNN*. The *smbCNN* model represents a custom-built shallow CNN trained using *TrIdent*'s native images, and *TrIdent*[*IRV2, SS*] represents an alternate *TrIdent*[*IRV2*] architecture trained using summary statistics based images (see *Viability of alternate architectures and methods* for details).

are also expected to be correlated. For methods that utilize summary statistics as input, the arrangement of haplotypes within a window does not affect the computation, as the statistics are derived from the aggregated properties of the window rather than the specific ordering of haplotypes.

Fine-Tuning vs. Feature Extraction

Building upon our comparison of five pretrained architectures in the previous subsection, we conducted an additional experiment to explore whether the strong performance of *TrIdent* is attributable to the model itself or to the pretrained feature extraction capabilities of the *InceptionResNetV2* backbone. We developed a modified model, termed *IRV2* (*InceptionResNetV2* architecture trained on *TrIdent* images), which extends the *InceptionResNetV2* architecture by incorporating a GAP layer and an output layer consisting of a single node with a sigmoid activation function but does not include an elastic-net style penalty for parameter fitting. The model was trained directly on the *TrIdent* input images, with starting weights obtained from pretrained weights of ImageNet.

The training dataset size for *IRV2* was the same as that utilized for the *TrIdent*[*IRV2*] model in one experiment (1,000 samples per class) and was increased to 10,000 samples per class in a second experiment. Following the training of *IRV2* on 1,000 samples per class, a slight enhancement in performance was noted in comparison to *TrIdent*[*IRV2*]. The accuracy improved by 1.7% on the CEU dataset (87.9% compared to 86.2%) and by 3.3% on the YRI dataset (90.85% compared to 87.55%; supplementary fig. S5,

Supplementary Material online). When trained on 10,000 samples per class, *IRV2* achieved even higher accuracies: 90.85% on CEU and 92.45% on YRI. These results illustrate the capacity of *InceptionResNetV2* to achieve superior performance with an increased number of training samples, consistent with observations in other domains where large-scale training datasets unlock additional classification accuracy and power (Yu et al. 2015; Alzubaidi et al. 2021).

The resource demands of training *IRV2* (see the *Computational resources and requirements* subsection of the *Methods*) underscore the computational challenges inherent in training deep neural networks from scratch using simulated data, rather than utilizing pretrained architectures. These findings highlight the balance between accuracy and computational efficiency. While *IRV2* outperformed *TrIdent*[*IRV2*], such gains come at the expense of significantly greater computational resources and training time. *TrIdent*[*IRV2*] demonstrates that by employing pretrained feature extraction, strong performance can be achieved with substantially fewer training samples and lower computational overhead. This balance positions *TrIdent*[*IRV2*] as a practical and effective solution for natural selection detection, especially where computational resources are constrained.

Performance in the Presence of Missing Genomic Regions

Due to undiscovered polymorphisms, haplotypic diversity can be reduced when segments of the genome are missing as a result of poor mappability, alignability, sequencing, or variant

calling (Alkan et al. 2011; Talkowski et al. 2011). As a consequence, this decline in local genomic variation can resemble a footprint of a selective sweep, misleading sweep detection approaches to mistakenly attribute such diversity as a positive selection in neutrally evolving genomic regions (Vitti et al. 2013; Mughal et al. 2020). Hence, we aim to evaluate whether *TrIdent*[*IRV2*] and *T-REx*, both with native and alternate ways of generating images, mistakenly identify neutrally evolving regions harboring missing genomic segments as selective sweeps as well as lose capacity to differentiate sweeps from neutrality under such settings. We removed polymorphisms from test replicates according to the human empirical distribution of CRG (Centre for Genomic Regulation) mappability and alignability scores (see *Filtration of empirical data* subsection of the *Materials and Methods*), which past studies have shown to negatively affect machine learning sweep classifiers (Mughal and DeGiorgio 2019; Mughal et al. 2020; Amin et al. 2023; Arnab et al. 2023).

We simulated additional sets of neutral and sweep test replicates under both CEU and YRI demographic histories using discoal, and introduced missing regions using the identical protocol as Arnab et al. (2023) based on the empirical human distribution. Specifically, we randomly chose one of the 22 human autosomes with probability proportion to the length of each autosome. From this autosome, we selected a starting genomic location uniformly at random for a 1.1 Mb segment. We removed single nucleotide polymorphisms (SNPs) that intersected with human genomic regions of low mappability and alignability (see *Filtration of empirical data and empirical image generation* subsection of the *Materials and Methods*). Because *TrIdent* only considers the central 499 SNPs when generating images from simulated replicates, we ensured that missing data blocks removed at least one of the central 499 SNPs, repeating the procedure until this criterion was satisfied. We found that this procedure removed on average approximately 23.1% of the original set of central 499 SNPs across all test replicates.

On the CEU test case, when *TrIdent*[*IRV2*] trained with its stock image generation technique is applied to replicates with missing polymorphisms, the overall accuracy drops by 4.45% (supplementary fig. S6, Supplementary Material online) compared to no missing segments (Fig. 3), whereas missing data causes *TrIdent*[*IRV2, alt*] trained on the alternate image generation technique to experience a drop in overall accuracy by 6.1% (supplementary fig. S6, Supplementary Material online) compared to no missing data (Fig. 3). In particular, while neutral detection rate was increased by roughly 1% with either style of input image, sweep detection rate was drastically decreased with missing segments regardless of image generation style (Fig. 3 and supplementary fig. S6, Supplementary Material online). While *TrIdent*[*IRV2*] saw a sweep detection rate drop of 10.5%, *TrIdent*[*IRV2, alt*] trained on the alternate image generation style saw an even higher drop of 13%. Conversely, when *T-REx* trained with its stock image generation technique is applied to test replicates with missing segments, its overall accuracy dropped by 3.65% (supplementary fig. S6, Supplementary Material online) compared to no missing polymorphisms (Fig. 3). On the other hand, *T-REx*[*alt*] trained on the alternate image generation style undergoes an 11.3% decrease in accuracy (supplementary fig. S6, Supplementary Material online) compared to no missing segments (Fig. 3). The neutral detection rate is preserved for both image generation styles, though the sweep detection rate for *T-REx*[*alt*]

showcases a significantly greater drop of 22.7% compared to the 7.1% decrease in classification accuracy for *T-REx* when both are applied to missing polymorphisms (Fig. 3 and supplementary fig. S6, Supplementary Material online).

On the YRI test case, however, the drop in test accuracies of all four models when presented with replicates containing missing genomic regions (supplementary fig. S7, Supplementary Material online) appears to be less severe than in the CEU test case (supplementary fig. S6, Supplementary Material online). Our results show a loss of 4.05% by *TrIdent*[*IRV2*], 2.65% by *TrIdent*[*IRV2, alt*], 2.80% by *T-REx*, and 7.50% by *T-REx*[*alt*] (supplementary fig. S7, Supplementary Material online) compared to test set with no missing data (Fig. 4). All four models present comparable neutral classification rates (from 84.3% to 88.5%). On the other hand, with a sweep detection rate of 81.7%, *TrIdent*[*IRV2*] again establishes itself as the most robust against missing genomic regions. The overall test accuracies lend further support to our observation that the *TrIdent* models, with accuracies of 83.50% and 82.65%, comfortably outperform the *T-REx* models, with accuracies of 80.0% and 76.35% (supplementary fig. S7, Supplementary Material online). Our findings demonstrate that the drops in accuracies for the *TrIdent* and *T-REx* model variants under missing genomic segments is generally driven by misclassification of sweep replicates as neutral, rather than the false attribution of selection for neutral settings.

Robustness Against Background Selection

Background selection poses a potential challenge for *TrIdent*[*IRV2*], as it has the potential to mislead sweep classifiers into falsely detecting signatures of positive selection. The occurrence of this phenomenon is caused by the elimination of harmful genetic variations through negative selection, resulting in distortions in the distribution of allele frequencies that may resemble positive selection (Charlesworth et al. 1993; Hudson and Kaplan 1995; Charlesworth 2012). Specifically, background selection can produce patterns in allele frequency distributions that resemble those caused by selective sweeps (Charlesworth et al. 1993, 1995, 1997; Keinan and Reich 2010; Seger et al. 2010; Nicolaisen and Desai 2013), which can result in misclassification of evolutionary events (Huber et al. 2016). However, recent studies suggest that selective sweeps and background selection leave distinct genetic footprints, particularly in haplotype distributions indicating that background selection is unlikely to pose a significant issue when analyzing haplotype data (Fagny et al. 2014; Schrider 2020; Lauterbur et al. 2023). Nevertheless, it is imperative to evaluate the resilience of *TrIdent*[*IRV2*] to the pervasive force of background selection that shapes patterns of genomic variation (McVicker et al. 2009; Comeron 2014).

We generated 1,000 test replicates using the forward-time simulator SLiM (Haller and Messer 2019) under the same genetic and demographic parameters of the CEU and YRI simulations used to train *TrIdent*, with the addition of background selection (Tennessen et al. 2012; Adrion et al. 2020). Specifically, in a simulated 1.1 Mb region, a 55 kb protein-coding gene was subjected to negative selection according to the methodology described by Cheng et al. (2017). This gene was composed of 50 exons each of length 100 bases, 49 introns each of length 1 kb, a 5′ untranslated region (UTR) of length 200 bases, and a 3′ UTR of length 800 bases, which are fairly close to mean lengths for these elements in

human genomes (Mignone et al. 2002; Sakharkar et al. 2004). Recessive ($h = 0.1$) deleterious mutations with selection coefficients (s) drawn from a gamma distribution with mean of -0.0294 and shape parameter of 0.184 (Boyko et al. 2008; Schrider and Kern 2017) arose within this gene, where 75%, 10%, and 50% of mutations in exons, introns, and UTRs, respectively, were deleterious. Consequently, we used the procedure outlined in the *Image Generation* subsection of the *Materials and Methods* to create input images, which we subsequently fed into trained *TrIdent*[*IRV2*] models as test observations.

Our findings suggest that the *TrIdent*[*IRV2*] models trained on the CEU and YRI datasets are resilient to background selection (supplementary fig. S8, Supplementary Material online), with the incidence of false sweep signals attributed to background selection closely paralleling the FPR obtained from neutral test observations (supplementary fig. S8, Supplementary Material online), particularly at acceptable FPRs (supplementary fig. S8, Supplementary Material online; bottom panels). This robustness to background selection is not only observed for settings in which recombination rates are drawn from the same distribution as the training sets (supplementary fig. S8, Supplementary Material online, left panels), but also for the challenging setting in which mean recombination rates are an order of magnitude smaller than what was used in the training sets (supplementary fig. S8, Supplementary Material online, right panels). Furthermore, for a mean recombination rate of 10^{-8}, *TrIdent*[*IRV2*] models identified 92.15% and 94% of background selection simulations as neutral for the CEU and YRI datasets, respectively. These rates are higher than the neutral detection rates of 87.1% and 87.7% for respective CEU and YRI neutral simulations (Figs. 3 and 4), indicating that *TrIdent*[*IRV2*] is even less likely to falsely attribute background selection as a sweep than it is for neutrally evolving regions.

Performance with Unphased Multilocus Genotypes

Sweep classifiers can face challenges when applied to unphased multilocus genotype data, which harbor less information than phased haplotypes, as selection acts to alter frequencies of nearby neutral haplotypic variation and only indirectly on multilocus genotype variation. Using unphased genotypes also decreases statistical power and introduces complexity in interpreting genetic signals, potentially omitting subtle yet significant signatures of selection within populations. However, *TrIdent* models are trained with phased haplotypic data, which is often difficult or impossible to reliably generate for many study systems, especially for most nonmodel organisms. Thus, to enhance the versatility of these models, it is crucial that they can accommodate unphased data. Given the demonstrated capability to detect sweeps using unphased multilocus genotypes in prior studies (Kern and Schrider 2018; Mughal and DeGiorgio 2019; Harris and DeGiorgio 2020; Gower et al. 2021; Arnab et al. 2023), we expect that *TrIdent* will continue to achieve excellent classification accuracy and power when applied to unphased data.

To generate *TrIdent* input images from unphased data, we first merge pairs of rows (haplotypes) into a single multilocus genotype to create values of zero, one, and two representing the number of copies of the minor allele at a diploid genotype. We then followed the same procedure described in the *Image generation* subsection of the *Materials and Methods* to create images for the *TrIdent*[*IRV2*] model. We refer to the

TrIdent[*IRV2*] model using images from unphased multilocus genotype data as *TrIdent*[*IRV2*, *MLG*]. We find that for both the CEU and YRI (supplementary fig. S9, Supplementary Material online) test cases, *TrIdent*[*IRV2*, *MLG*] achieves comparable, though marginally lower, accuracies and powers to *TrIdent*[*IRV2*], with a 1.95% drop in classification accuracy on the CEU test case and a more narrow drop of 0.15% in accuracy on the YRI test case. Moreover, under both scenarios, *TrIdent*[*IRV2*, *MLG*] is more conservative than *TrIdent*[*IRV2*], with a slight bias toward predicting observations as neutral (supplementary fig. S9, Supplementary Material online). Overall, these results indicate that while *TrIdent*[*IRV2*, *MLG*] is slightly less accurate than *TrIdent*[*IRV2*], it remains a robust method for analyzing unphased multilocus genotype data, demonstrating the flexibility of the *TrIdent* approach.

Capacity to Uncover Incomplete Sweeps

The less pronounced genomic footprints of incomplete or partial sweeps may increase false negative rates of sweep classifiers by misleadingly assigning such patterns as neutral (Schrider et al. 2015; Xue et al. 2021). Specifically, these genomic signals are less pronounced than what is expected by recent complete sweeps, including localized and weaker linkage disequilibrium, shorter and less frequent haplotypes, a more mildly distorted site frequency spectrum, and a comparatively more marginal decrease in genetic variation (Vy and Kim 2015). As a result, it is important to evaluate the power of *TrIdent* to detect incomplete sweeps relative to other sweep classifiers. To evaluate power under this setting, we used `discoal` (Kern and Schrider 2016) to generate sweep test replicates for both the CEU and YRI demographic histories for which an advantageous allele does not reach fixation. Specifically, we simulated 1,000 replicates for each incomplete sweep scenario, considered situations for which the beneficial allele stopped being advantageous at a frequency of $f_{end} \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$ while fixing all other genetic, demographic, and selection parameters as detailed within the *Simulation protocol* subsection of the *Methods*. We then applied *TrIdent*[*IRV2*] and *T-REx* originally trained on complete sweeps to these incomplete sweep test sets.

We evaluated both accuracy and power (true positive rate) at a 5% FPR to detect incomplete sweeps (supplementary fig. S10, Supplementary Material online). Under the CEU demographic history, accuracy for both *TrIdent*[*IRV2*] and *T-REx* climbed with increasing final frequency of the beneficial allele f_{end}, with values as low as 31.06% and 21.7% at $f_{end} = 0.5$ for *TrIdent*[*IRV2*] and *T-REx*, respectively, and achieving values of 83.15% and 70.1% at $f_{end} = 0.9$ for *TrIdent*[*IRV2*] and *T-REx*, respectively. Notably, *TrIdent*[*IRV2*] achieves an accuracy close to 80% for incomplete sweeps to a frequency of $f_{end} = 0.8$, whereas *T-REx* never reaches 80% sweep detection rate even for complete sweeps for which $f_{end} = 1.0$ (supplementary fig. S10, Supplementary Material online). A similar increasing trend in power as a function of the degree of sweep completeness can be observed (supplementary fig. S10, Supplementary Material online), with values as low as 0.201 and 0.11 at $f_{end} = 0.5$ for *TrIdent*[*IRV2*] and *T-REx*, respectively, and achieving values of 0.73 and 0.56 at $f_{end} = 0.9$ for *TrIdent*[*IRV2*] and *T-REx*, respectively. Moreover, *TrIdent*[*IRV2*] achieves a power slightly above 0.7 at $f_{end} = 0.8$ (which is virtually identical to the power for a complete sweep), whereas *T-REx* never even achieves a power of 0.6 for complete sweeps.
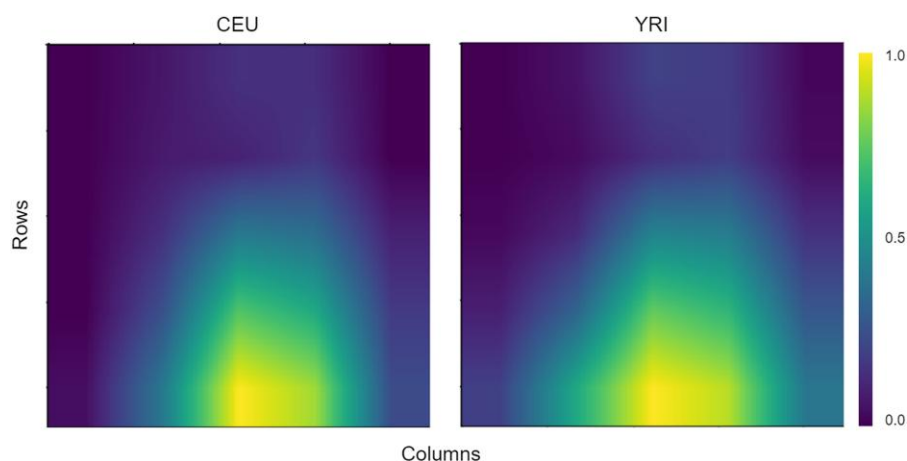
Fig. 7. Heatmaps of mean GradCAM from *InceptionResNetV2* applied to CEU (left panel) and YRI (right panel) training data, with the mean taken across 2,000 training observations with 1,000 observations from each of the neutral and sweep classes.

Thus, *TrIdent*[*IRV2*] consistently and significantly outclassed *T-REx* in terms of accuracy and power to detect incomplete sweeps.

For the YRI test case, we observed an overall higher accuracy and power for detecting incomplete sweeps compared to the CEU test case. Specifically, the accuracy for both *TrIdent*[*IRV2*] and *T-REx* increased with f_{end}. At $f_{end} = 0.5$, the accuracy values for *TrIdent*[*IRV2*] and *T-REx* were 44.11% and 38.6%, respectively, and climbed to 87.03% for *TrIdent*[*IRV2*] and 78.99% for *T-REx* at $f_{end} = 0.9$. A similar trend was seen in power, with values starting from 0.307 and 0.212 at $f_{end} = 0.5$ for *TrIdent*[*IRV2*] and *T-REx*, respectively, and respectively rising to 0.7404 and 0.607 at $f_{end} = 0.9$. These results show that *TrIdent*[*IRV2*] consistently and significantly outperformed *T-REx* in both accuracy and power for the YRI test case, expectedly surpassing the performance levels observed in the CEU test case that includes a recent, severe population bottleneck. In general, the results highlight the ability of *TrIdent* in recognizing even modest adaptive signals in genomes.

Interpretability of the Sweep Classifier

In nonnatural image classification tasks, class activation maps can reveal the decision-making process of a trained model (Zhai and Shah 2006; Zhou et al. 2016). In particular, these maps are able to highlight regions of pixels in nonnatural images, like those generated here to discriminate sweeps from neutrality, where patterns may not be immediately discernible to humans that explain or validate what features the model places emphasis when classifying observations. In addition, class activation maps can aid in model selection based on performance and robustness in image classification tasks by acting as a diagnostic tool to identify errors in the learning process of a model by visualizing key regions of input images, allowing for targeted adjustments to improve accuracy and generalization and thereby creating more transparent and reliable classification systems (Zhou et al. 2016).

For pretrained CNN models, gradient-weighted class activation mapping (GradCAM) provides a mechanism for creating class activation maps that are easy to understand and work with (Selvaraju et al. 2017). To accomplish this task, GradCAM makes use of the gradient information that flows into the last convolutional layer of the CNN during its backward pass. It computes the importance of each feature map by taking the gradients of the target class score with respect to these feature maps and weighting them accordingly. By highlighting the relative importance of each pixel, this approach successfully pinpoints the regions within an image that help discriminate among classes (Selvaraju et al. 2017).

We employed GradCAM to generate class activation maps from each of the training set images based on their output values from the last convolution layer of *TrIdent*[*IRV2*]. The resulting maps for both the CEU and YRI test cases revealed a concentrated focus in the lower-middle region of the images (Fig. 7), coinciding with regions of input images that might distinguish sweeps from neutrality on average (Fig. 2). Moreover, this consistent regional focus across both CEU and YRI test cases suggests that *TrIdent*[*IRV2*] effectively captures the critical haplotype features driving the divergence between sweep and neutrality in these populations.

Ability to Predict Sweep Parameters

So far, we have assessed the sweep detection accuracy and power of *TrIdent*[*IRV2*] in relation to the sweep classifiers *T-REx*, diploS/HIC and many architectural variations of the the *TrIdent* model. We next consider how features obtained from transfer learning architectures fare on regression tasks for predicting evolutionary parameters of sweeps. We trained three distinct models, each tasked with inferring one of the three selection parameters used to generate sweep replicates for training the *TrIdent*[*IRV2*] classifier. These parameters are: the number of generations in the past when the beneficial allele reached fixation (τ), the selection strength acting on the beneficial allele (s), and the frequency of the beneficial allele at the onset of selection (f). As with the *TrIdent* classifiers, we first extracted GAP layer outputs of *InceptionResNetV2* from the 1,000 sweep replicates to represent the set of features in the training dataset. As the output for each regression model, we performed a logarithmic transformation of the parameters so that they take both positive and negative values, rather than only nonnegative values, as well as to better highlight parameters that were drawn across different orders of magnitude (see *Modeling description* subsection).

Our initial evaluation of a linear model to predict the selection parameters showed that such a model was not accurate enough based solely on the *InceptionResNetV2* extracted
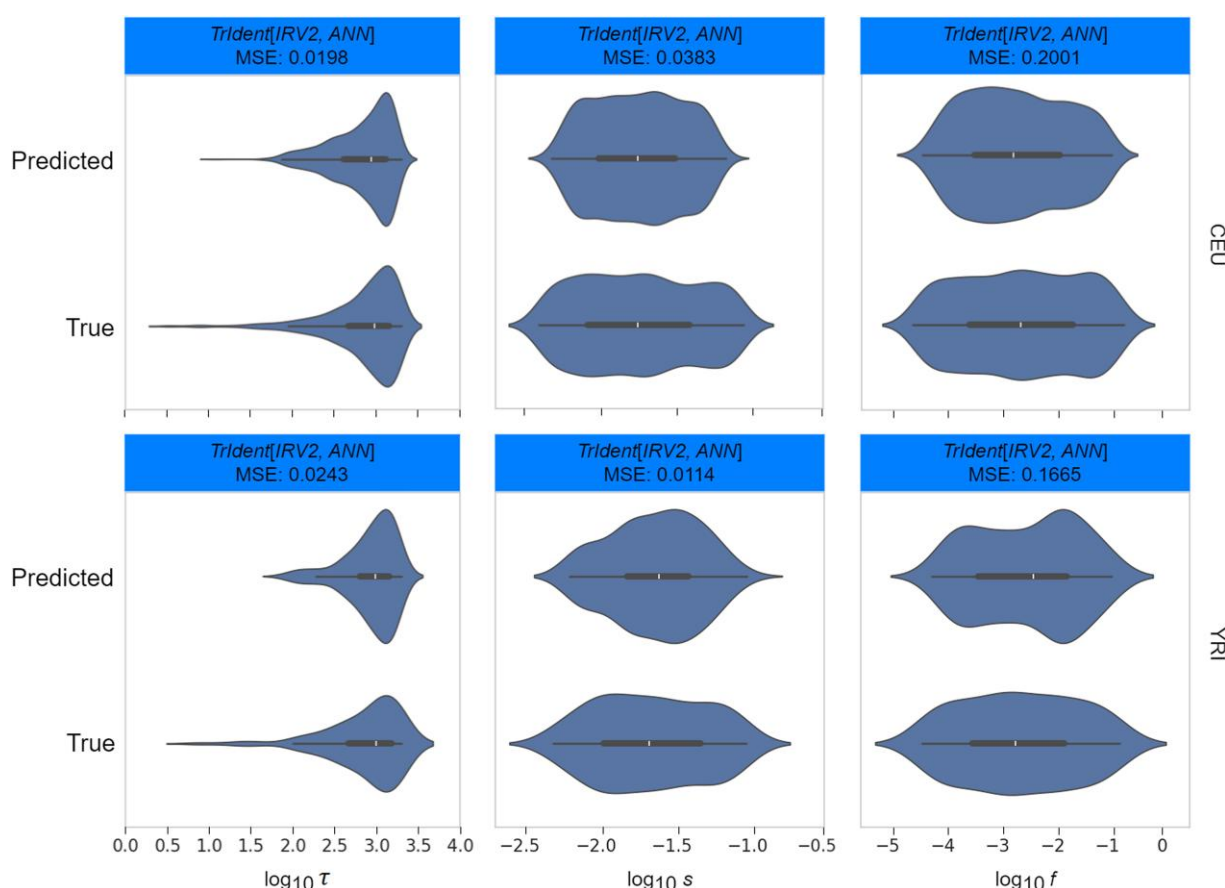
Fig. 8. Summaries of distributions for true and predicted values of selection parameters (*f*, *s*, and *τ*) using the nonlinear *TrIdent*[*IRV2, ANN*] regression model. Distributions are summarized using violin plots with embedded box plots for the CEU (top) and YRI (bottom) datasets.

features. We therefore chose to employ nonlinear models based on ANNs for predicting sweep parameters, and we term this nonlinear regression model *TrIdent*[*IRV2, ANN*], similar to the nonlinear classification model of the *Viability of alternate architectures and methods* subsection. To determine the architecture of each sweep parameter predictor model (*f*, *s*, and *τ*), we use a model selection scheme identical to that of the nonlinear *TrIdent*[*IRV2, ANN*] classifier, with the exception of assessing model fit with MSE loss instead of binary cross entropy along with the usage of linear activation in the output layer compared to sigmoid activation. Here, we choose the best number of hidden layers, number of nodes in each layer, and activation function used in each layer for each of the three regression models based on smallest validation MSE.

To compare distributions between true and predicted parameter values, we summarize the distributions using violin plots to capture distribution shapes with embedded box plots to depict distribution locations and spreads (Fig. 8). For *τ*, the medians for true and predicted distributions are comparable in both CEU and YRI populations. The MSE is 0.0198 for CEU and 0.0243 for YRI, and the distributions are left skewed, with the difference between the second and first quartiles of the predicted distribution smaller in the YRI than in CEU. For *s*, the MSE is 0.0383 for CEU and 0.0114 for YRI, with the predicted distributions preserving the characteristic "lip" shape observed in the true distribution, though with predicted distributions tighter than the true distributions. The medians of the true and predicted distributions are comparable in both datasets, but the inter-quartile range (IQR) is narrower

in the predicted distributions. For *f*, the predicted distributions display skewed "lip" shapes—right-skewed for CEU and left-skewed for YRI—though the true distributions show a more symmetrical "lip" shape. Moreover, the predicted median is shifted slightly downward relative to the true median in the CEU, whereas it is shifted upward in the YRI. However, whereas the IQR is smaller for the predicted distribution in the CEU dataset than for the true distribution, the IQR for the predicted distribution in the YRI dataset is comparable to that of the true distribution, with MSE values of 0.2001 and 0.1665 for CEU and YRI, respectively. While the MSE values may appear high, it is important to note that the parameters were logarithmically scaled, which stretches their bounds, inherently inflating the apparent magnitude of the MSE compared to unscaled values. Overall, in terms of MSE, *TrIdent*[*IRV2, ANN*] seems to perform better on predicting *s* and *f* on the YRI dataset compared to on the CEU dataset, and *τ* on the CEU dataset compared to on the YRI dataset.

To benchmark against a comparative baseline for selection parameter estimates, we employed the selection coefficient inference tool CLUES2 (Vaughn and Nielsen 2024). While CLUES2 is limited to estimating the selection coefficient, one of the three parameter estimates by the *TrIdent*[*IRV2, ANN*] regression model, it serves as a useful reference point. Details of our CLUES2 pipeline are presented in the *Application of CLUES2* subsection of the *Materials and Methods*. We generated ancestral recombination graphs (ARGs) using the entire 1.1 Mb simulated replicates with

SINGER (Deng et al. 2024), as it is one of the two software supported by CLUES2. We implemented CLUES2 as closely to the recommended settings as possible, given the constraints imposed by our simulated replicates. In both CEU and YRI test cases, CLUES2 lagged considerably behind the accuracy and precision of *TrIdent*[*IRV2*, *ANN*] (supplementary fig. S11, Supplementary Material online). To maintain consistency with *TrIdent*[*IRV2*, *ANN*], we applied a logarithmic transformation to the selection coefficient estimates from CLUES2, which resulted in MSEs on this transformed scale of 0.454 for CEU and 0.402 for YRI.

Application to Human Genomic Data

Given the excellent performance of *TrIdent*[*IRV2*], we next applied it to variant calls and phased haplotypes of 99 individuals in the CEU population and 108 individuals in the YRI population from the 1000 Genomes Project (The 1000 Genomes Project Consortium 2015). This application served as a proof-of-concept, where we evaluate the ability of *TrIdent* to recapitulate sweep candidates established in the literature from European and African humans, as well as to potentially discover novel candidates.

Before scanning the CEU and YRI genomes, we eliminated SNPs in specific regions using the procedure detailed in the *Filtration of Empirical Data* subsection of the *Materials and Methods*. Following the elimination of these SNPs, we selected the initial 499 SNPs of a chromosome and created a 299×299 input image of haplotype variation (see *Modeling description* subsection of the *Results*). From this starting image, we created new empirical input images by advancing the 499-SNP window by a stride of 10 SNPs along the chromosome, repeating the process for all 22 autosomes. These images are then fed through the *TrIdent*[*IRV2*] pretrained model, and the GAP layer outputs are used as input to the trained logistic regression classifier (see *Modeling description* subsection of the *Results*).

Table 1 Percentage of windows classified as a sweep by *TrIdent* under sweep probability threshold of 0.9 for each of the autosomes within the CEU and YRI 1000 Genomes Project populations

Chromosome	CEU	YRI
1	5.48	4.84
2	4.68	3.78
3	4.02	2.16
4	4.65	6.11
5	4.22	4.64
6	4.47	5.06
7	4.38	4.40
8	5.92	1.92
9	5.15	4.51
10	3.96	3.77
11	3.75	4.40
12	4.76	3.86
13	3.89	3.66
14	5.32	3.78
15	5.10	3.52
16	5.53	5.14
17	6.67	3.32
18	4.34	2.01
19	7.08	6.69
20	6.17	3.06
21	4.97	4.07
22	7.17	2.65

In the CEU scan, we found that the majority of genomic windows are classified as neutral with a probability threshold of 0.9 for calling sweeps (approximately 95.18%; Table 1). The threshold of 0.9 for individual windows is strategically chosen to reduce the risk of false positives, thus prioritizing the most significant candidate sweeps and improving the reliability of the detection process. We evaluated the false positive and true positive rates at this threshold of 0.9 using the CEU and YRI simulated test sets (supplementary fig. S12, Supplementary Material online). This stringent threshold demonstrated its effectiveness in minimizing false positives, achieving high detection rates for neutral replicates (97.5% for CEU and 97.7% for YRI, compared to 87.1% for CEU and 87.7% for YRI with a threshold of 0.5). While the detection rates for sweep replicates were notably lower (69.9% for CEU and 72.2% for YRI, compared to 85.3% for CEU and 87.4% for YRI with a threshold of 0.5), this reflects a more conservative approach to identifying true adaptive signals.

Additionally, we require a minimum mean prediction probability of 0.9 across 10 consecutive prediction windows for sweep footprint detection. This stringent threshold is designed to filter out potentially spurious observations with high sweep probability, ensuring that only the most robust signals are identified (Arnab et al. 2023). This approach resulted in 1,206 candidate sweep peaks, which intersected 575 genes. Out of these genes, *LCT* shows a clear peak with a 10-window mean peak probability score of 0.93 (Fig. 9a). The detection of *LCT* serves as a positive control, because it has been identified as a recent sweep candidate in numerous studies (Tishkoff et al. 2007; Field et al. 2016; Ségurel and Bon 2017) with strong estimates of selection pressure (Bersaglieri et al. 2004; Gerbault et al. 2009). The *LCT* gene codes for the lactase enzyme that hydrolyzes lactose, a disaccharide in milk and dairy products. Early agriculture and dairy farming in Europe profoundly influenced lactase persistence. Domesticating animals for milk grew more common when hunter-gatherer societies became sedentary agricultural groups. We also applied the *TrIdent* regression model to predict selection parameters at *LCT* (Table 2), with estimates that a sweep on a standing genetic variant at frequency $f \approx 0.01$ became beneficial with strength $s \approx 0.06$ and completed $\tau \approx 46$ generations ago (approximately 1,300 years ago assuming a generation time of 29 years). For comparison, Bersaglieri et al. (2004) estimated that the persistence-associated haplotype began to increase rapidly in frequency between 2,188 and 20,650 years ago from selection with coefficients ranging from 0.09 to 0.19 across different European populations at this locus. Thus, the estimates produced by *TrIdent* seem to generally agree well with the findings of Bersaglieri et al. (2004).

Moreover, the *MCM6* gene located upstream of *LCT* not only exceeds the mean prediction probability threshold of 0.9 but also has a higher probability compared to *LCT* (Fig. 9a). Selection parameter predictions at this gene (Table 2) support a sweep on a standing genetic variant at frequency $f \approx 0.1$ became beneficial with strength $s \approx 0.04$ and completed $\tau \approx 48$ generations or approximately 1,400 years ago. Previous studies also detected strong selection signals at *MCM6* within Europeans, though their selection coefficient estimates were generally lower (0.0161 in Stern et al. 2019 and 0.018 in Mathieson and Mathieson 2018). The time frame that selection ended is similar to that of *LCT*, but with a softer sweep of lower strength. Due to its regulatory control of the expression of *LCT* (Labrie et al. 2016; Anguita-Ruiz et al.
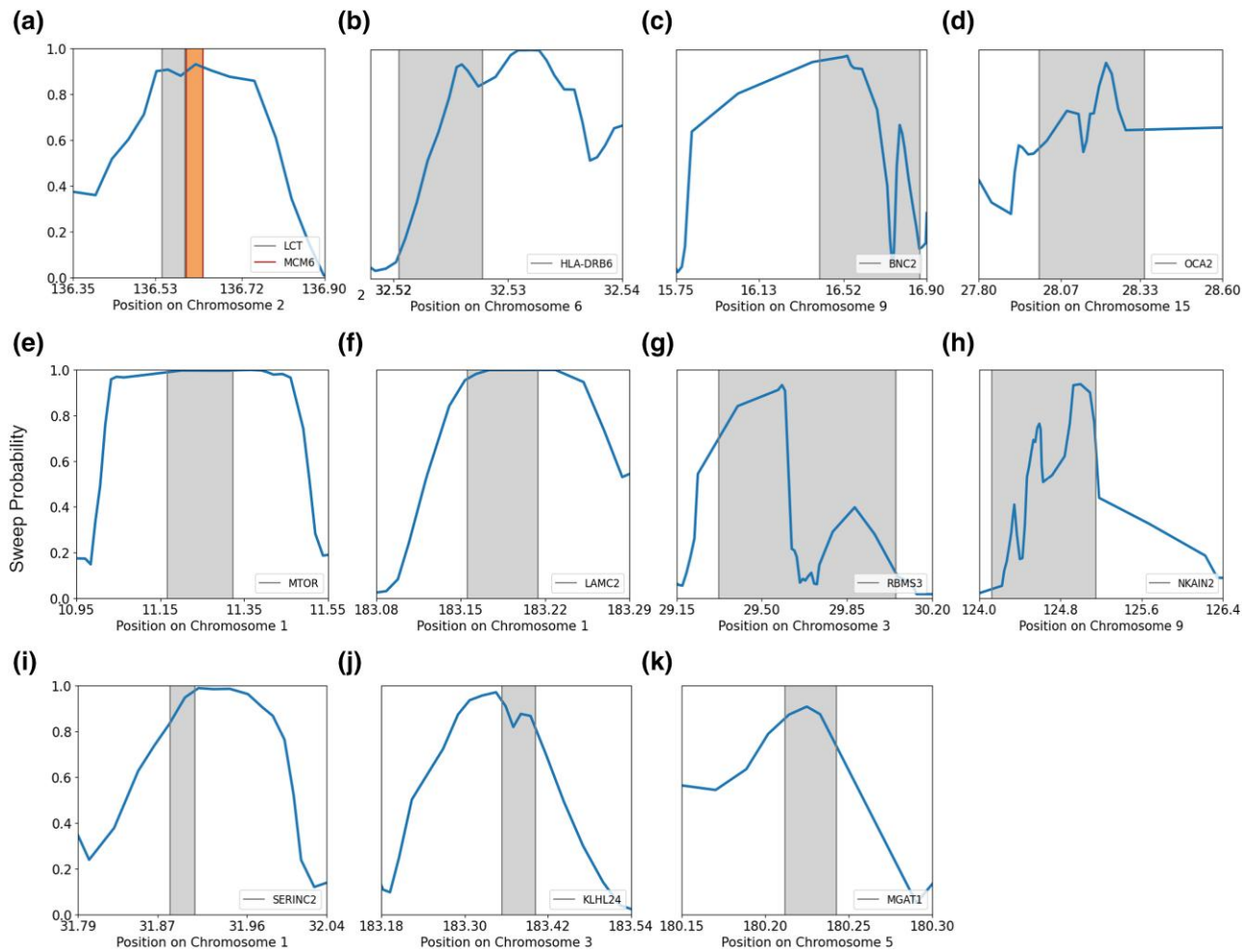
Fig. 9. Identified candidate sweep regions from the genome-wide scan produced using the trained *TrIdent*[*IRV2*] model on the central European humans (CEU) population in the 1000 Genomes Project dataset. Regions were classified as being under positive selection if they had 10 consecutive windows with a sweep probability higher than 0.9. A total of 575 genes across 22 autosomes exhibit qualifying signs of selective sweeps, of which a few of the most interesting candidates are reported here (a–k). supplementary Figure S21, Supplementary Material online provides a visual representation of the haplotype diversity surrounding the candidate genes in the plotted panels.

Table 2 *TrIdent* regression model inferences for the frequency at which a selected allele became beneficial (f), selection strength (s), and time at which the beneficial allele reached fixation (τ) for genes reported in Fig. 9 or the CEU scan

Chromosome	f	s	τ (generations)	Genes
1	0.0884	0.0700	1295.7883	*MTOR*
1	0.0601	0.0814	156.0666	*SERINC2*
1	0.0252	0.0505	228.1717	*LAMC2*
2	0.0122	0.0623	46.1186	*LCT*
2	0.0962	0.0373	48.2734	*MCM6*
3	0.0016	0.0133	1340.9021	*RBMS3*
3	0.0107	0.0321	842.0184	*KLHL24*
5	0.0490	0.0257	753.2627	*MGAT1*
6	0.0399	0.1051	982.6476	*HLA-DRB6*
9	0.0550	0.0191	274.6670	*BNC2*
9	0.0068	0.0163	1141.4152	*NKAIN2*
15	0.0229	0.0598	15.0190	*OCA2*

2020), *MCM6* also serves a positive control. An enhancer within one of the introns of the *MCM6* gene has been discovered to impact the production of the lactase enzyme (Anguita-Ruiz et al. 2020), and has been found to show sweep

signals by other studies (Oleksyk et al. 2010; Cheng et al. 2017; Amin et al. 2023).

One of the most pronounced sweep candidates identified by *TrIdent* is *OCA2*, with a 10-window mean sweep probability of 0.97 (Fig. 9d), estimated to have completed $\tau \approx 15$ generations ago (Table 2), indicating that a putative sweep at *OCA2* completed more recently compared to that of *LCT* or *MCM6*. Previous studies have found *OCA2* as a target of positive selection in Europeans (Voight et al. 2006; Sulem et al. 2007; Wilde et al. 2014; Mughal et al. 2020), and therefore this gene also serves as an additional positive control. We estimate a selection coefficient of $s = 0.059$ for *OCA2*, indicating strong selection at this locus. For comparison, Stern et al. (2019) also identified evidence of selection at *OCA2*, estimating the strength at $s = 0.04$, whereas (Mughal et al. 2020) reported $s = 0.06$ for the same locus. Hence, our selection strength estimates are consistent with prior estimates at this gene. Eye color is associated with variants within the *OCA2* gene, which plays a crucial role in the production and dispersion of melanin, the pigment that gives color to hair, skin, and eyes (Duffy et al. 2007). The amount and type of iris melanin, which determines eye color, can vary due to structural variations in *OCA2* (Sturm and Larsson 2009).

The human leukocyte antigen (HLA) system serves as a compelling example of how natural selection has operated to preserve genetic diversity. HLA genes, particularly those in class I and class II, play a crucial role in presenting antigens to T cells, which are responsible for initiating cell-mediated immune responses (Shankarkumar 2004). The extensive range of alleles and notable polymorphisms observed in HLA class I and class II genes (Hedrick and Thomson 1983) serves as evidence of the ongoing selection that promotes genetic diversity at these immune-related locations. *TrIdent* follows the recent trend of sweep detectors (Goeury et al. 2018; Kern and Schrider 2018; Harris and DeGiorgio 2020; Amin et al. 2023; Arnab et al. 2023) in uncovering genes within the HLA region. Notably, we found that the class II HLA gene *HLA-DRB6* shows a clear signal consistent with positive selection, attaining a 10-window mean sweep probability of 0.92 (Fig. 9b). Other studies have identified sweep signals at this gene in Europeans (DeGiorgio and Szpiech 2022; Arnab et al. 2023), as well as other class I and class II genes within the region (Albrechtsen et al. 2010; Goeury et al. 2018; Harris and DeGiorgio 2020; DeGiorgio and Szpiech 2022), lending credence to this finding. Furthermore, *TrIdent* predicts a notably high selection coefficient of $s = 0.1051$ for *HLA-DRB6*. Mughal et al. (2020) have also reported high selection coefficients for an HLA gene, finding $s = 0.14$ at *HLA-DRB1*, further supporting the notion that genes in the HLA region have been subject to intense selective pressures.

BNC2 is another well-supported sweep candidate with a peak mean sweep probability of 0.99 (Fig. 9c). This gene encodes a protein that plays a vital role in various essential cellular processes. These processes include controlling the expression of genes that code for proteins responsible for binding and regulating collagen, as well as for facilitating the growth of new tissues (Orang et al. 2023). Its involvement in key developmental pathways is supported by its association with skin cell differentiation (Jacobs et al. 2013). A structural variant in the first intron of *BNC2* causes lighter skin color by reducing *BNC2* expression in human melanocytes (Visser et al. 2014; Szpak et al. 2019). In addition, some haplotypes at *BNC2* have been hypothesized to be introgressed from Neanderthals with high frequency that influences skin pigmentation levels (Visser et al. 2014; Szpak et al. 2019; McArthur et al. 2021). Prior selection scans suggest that this variation at this gene represents a mode of positive selection termed adaptive introgression (Racimo et al. 2015, 2017; Mughal et al. 2020; Gower et al. 2021), by which selection acts on variants residing on haplotypes that were donated from another species through introgression. Furthermore, the likely importance of *BNC2* in tumor growth has made it a focal point in cancer research (Cesaratto et al. 2016; Wu et al. 2016; Orang et al. 2023), with some cancers linked to alterations in *BNC2* expression and function (Wu et al. 2016; Orang et al. 2023).

Along with the established candidate *BNC2*, *TrIdent* also detected several other cancer-related genes as sweep candidates. Specifically, with a 10-window mean sweep probability of 0.996, we identified the *MTOR* gene (Fig. 9e), which produces the protein mTOR that is a critical regulator of cellular growth and survival and that is frequently dysregulated in breast cancer (Miricescu et al. 2020). Cancer progression and uncontrolled growth of cells are caused by abnormal mTOR activation (Takei and Nawa 2014; Costa et al. 2015). *RBMS3* is another detected sweep candidate

(10-window mean sweep probability of 0.9969; Fig. 9g) that is linked to tumor suppression in breast cancer (Yang et al. 2018). Another candidate gene *LAMC2* (10-window mean sweep probability of 0.9999; Fig. 9f), which encodes the gamma-2 subunit of the protein laminin, is commonly upregulated in oral cancer (Nguyen et al. 2017). Furthermore, the gene *NKAIN2* may inhibit tumors in some cancers (Zhao et al. 2015), which is another candidate identified by our scan with a 10-window mean sweep probability of 0.98 (Fig. 9h). A few other intriguing sweep candidates revealed by our scan are *SERINC2* (10-window mean sweep probability of 0.997; Fig. 9i), *KLHL24* (10-window mean sweep probability of 0.995; Fig. 9j), and *MGAT1* (10-window mean sweep probability of 0.92; Fig. 9k), with association studies linking variants in *SERINC2* with alcohol dependence in European women (Zuo et al. 2013, 2014, 2015), mutations in *KLHL24* associated with loss of keratin 14, which provides structural support to epithelial cells (Lin et al. 2016), and a role of *MGAT1* in the development of type 1 diabetes in Europeans (Rudman et al. 2023). There appears to be a trend in recent machine learning studies where genes associated with deleterious phenotypes, including cancer, are identified as candidate targets of positive selection (Schrider and Kern 2017; Mughal and DeGiorgio 2019; Amin et al. 2023; Arnab et al. 2023). Schrider and Kern (2017) suggested that weakly deleterious alleles may have hitchhiked to high frequency alongside beneficial variants, resulting in the present-day manifestation of problematic traits. This phenomenon could also be explained by the historical advantages conferred by these genes in past environments, which outweighed their negative impacts under contemporary conditions (Di Rienzo and Hudson 2005; Di Rienzo 2006). Such hitchhiking events illustrate the complexity of evolutionary processes and the tradeoffs between short-term benefits and long-term consequences in gene selection, with further work needed to fully understand these dynamics and their implications for modern human health.

In the YRI scan, we found that most genomic windows are classified as neutral with a probability threshold of 0.9 for calling sweeps (approximately 95.41%; Table 1). Using a 10-window mean sweep probability of 0.9, we identified 2,145 candidate sweep peaks, which intersected 666 genes. This scan revealed several prominent candidate selection signals at genes supporting some findings from the CEU scan and exposing distinct selection patterns particular to the YRI population. Several genes highlighted as possible adaptive targets in previous studies of the YRI that were also recovered by our scan include *NNT*, *HEMGN*, *SYT1*, *GRIK5*, and *APOL1* (Fig. 10) (Voight et al. 2006; Pickrell et al. 2009; Fagny et al. 2014; Pierron et al. 2014; Harris et al. 2018; Harris and DeGiorgio 2020; Mughal et al. 2020). We also applied the *TrIdent* regression model to predict selection parameters at reported sweep candidates found in the YRI scan (Table 3).

The *NNT* gene, located on chromosome 5, exhibits a clear signal with a 10-window mean sweep probability score of 0.989 (Fig. 10a). This gene codes for a protein named nicotinamide nucleotide transhydrogenase, which is vital for cellular energy metabolism (Yin et al. 2012; Xiao et al. 2018) and therefore may have possible connection to metabolic adaptations in the YRI population. The *HEMGN* gene on chromosome 9 is another noteworthy candidate, with a mean sweep probability score of 0.974 (Fig. 10c) that is involved in hematopoiesis by which the body forms blood cells (Li et al.
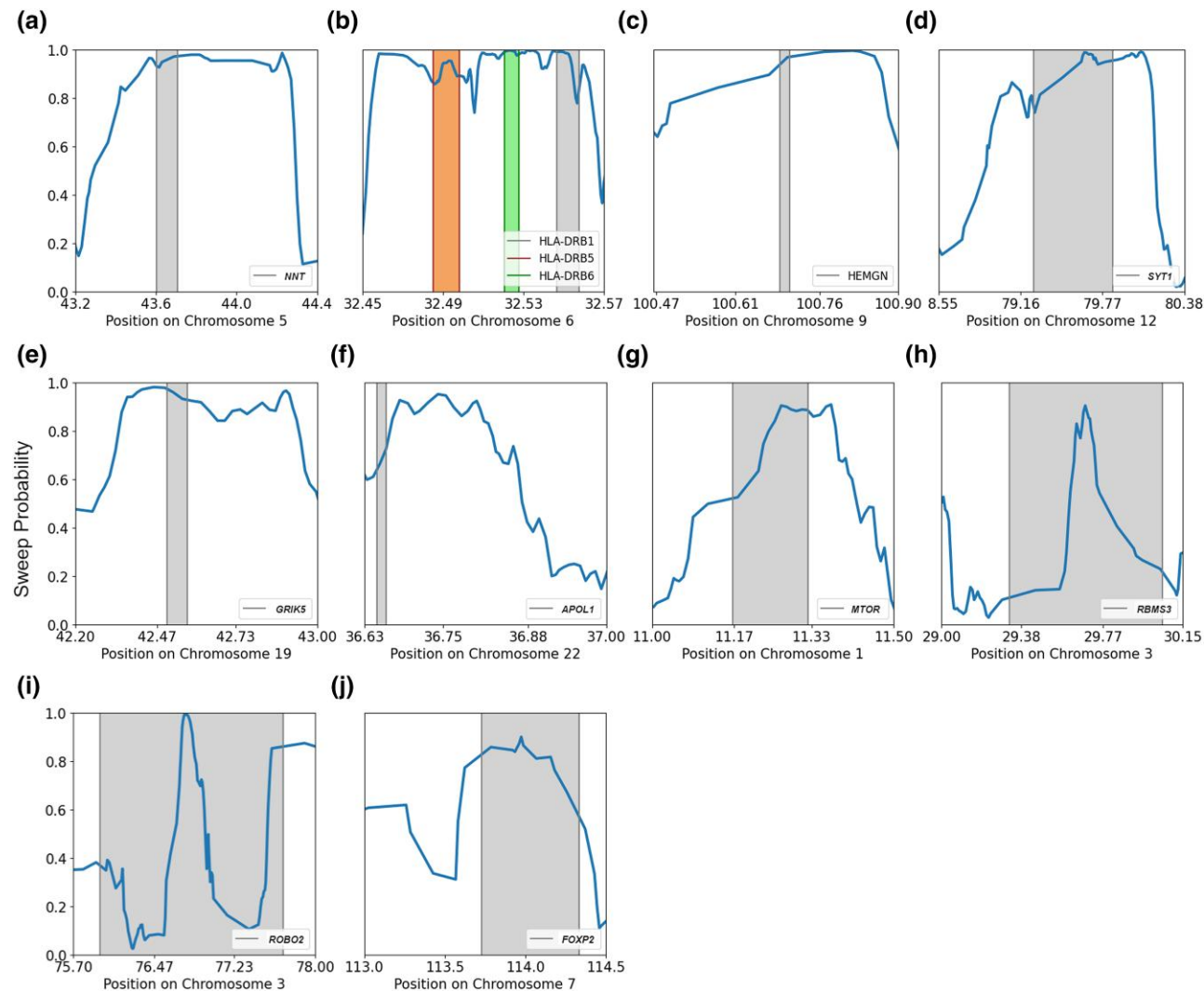
Fig. 10. Identified candidate sweep regions from the genome-wide produced using the trained *TrIdent*[*IRV2*] model on the sub-Saharan African (YRI) population in the 1000 Genomes Project dataset. Regions were classified as being under positive selection if they had ten consecutive windows with a sweep probability higher than 0.9. A total of 666 genes across 22 autosomes exhibit qualifying signs of selective sweeps, of which a few of the most interesting candidates are reported here (a–j). supplementary fig. S22, Supplementary Material online provides a visual representation of the haplotype diversity surrounding the candidate genes in the plotted panels.

Table 3 *TrIdent* regression model inferences for the frequency at which a selected allele became beneficial (*f*), selection strength (*s*), and time at which the beneficial allele reached fixation (*τ*) for genes reported in Fig. 9 or the YRI scan

Chromosome	f	s	τ (generations)	Genes
1	0.0418	0.0678	1049.3189	*MTOR*
3	0.0097	0.0879	89.8202	*RBMS3*
3	0.0387	0.1033	2710.1717	*ROBO2*
5	0.0290	0.0579	1546.6411	*NNT*
6	0.0832	0.0322	37.2414	*HLA-DRB1*
6	0.0133	0.0323	61.7333	*HLA-DRB5*
6	0.0063	0.0731	69.3420	*HLA-DRB6*
7	0.0011	0.0900	2901.9021	*FOXP2*
9	0.0200	0.0354	1883.3812	*HEMGN*
12	0.0222	0.0452	2024.0176	*SYT1*
19	0.0368	0.0478	675.6476	*GRIK5*
22	0.0880	0.0438	2000.9901	*APOL1*

2004; Jiang et al. 2010). Possible responses to environmental variables or pathogenic pressures may have prompted changes in blood cell formation and function. Another strong candidate highlighted by our scan with mean sweep probability of 0.987 is the *SYT1* gene on chromosome 12 (Fig. 10d). *SYT1* encodes Synaptotagmin 1, an essential regulator of neurotransmitter release in the nervous system (Xu et al. 2009; Park and Ryu 2018), and the adaptive relevance of this gene may be connected to cognitive and neurological processes in the YRI population. The *GRIK5* gene on chromosome 19, which encodes a subunit of kainate receptors involved in synaptic transmission and plasticity (Shibata et al. 2006), also shows a prominent signal with a mean sweep probability of 0.991 (Fig. 10e). Schrider and Kern (2017) also identified sweep signals at several glutamate receptor genes, highlighting the potential role of neurotransmitter receptors as targets of recent selection in humans. Furthermore, the gene *APOL1* on chromosome 22, which is associated with kidney diseases in individuals of African ancestry (Kruzel-Davila et al. 2017), has been detected with a relaxed qualification criteria by *TrIdent* with a mean sweep probability of 0.801 (Fig. 10f).

Within the HLA region of chromosome 6, the *HLA-DRB1*, *HLA-DRB5*, and *HLA-DRB6* genes show strong signals consistent with positive selection (Fig. 10b). In particular, the

HLA-DRB6 gene, also identified in our CEU scan, attains a 10-window mean sweep probability of 0.9989. Together with the discovery of noteworthy signals in two neighboring candidate genes (*HLA-DRB1* with a mean sweep probability of 0.963 and *HLA-DRB5* with a mean sweep probability of 0.988), these findings emphasize the ongoing selection pressures to maintain genetic diversity in this chromosomal region. Though *TrIdent* was primarily trained to detect selective sweeps that reduce haplotypic diversity, regions subject to recent balancing selection can sometimes exhibit signatures resembling those of soft or incomplete sweeps (Hermisson and Pennings 2005; Fijarczyk and Babik 2015). This similarity is because balancing selection, particularly when recent or strong, may lead to localized increases in linkage disequilibrium and allele frequency patterns that mimic the sweep-like signals captured by the model.

The signals found by *TrIdent* at *RBMS3* on chromosome 3 and *MTOR* on chromosome 1, identified in both YRI and CEU scans, underscore their broad adaptive significance. The relatively lower mean sweep probabilities in the YRI scan (0.911 for *RBMS3* and 0.906 for *MTOR*; Figs. 10g and 10h) compared to CEU may reflect differences in sweep completion status at these genes between the populations. Additionally, two particularly interesting candidates in the YRI genome are *ROBO2* (mean sweep probability score of 0.9975) on chromosome 3 (Fig. 10i) and *FOXP2* (mean sweep probability score of 0.916) on chromosome 7 (Fig. 10j). Though *FOXP2* is a classical candidate for selection related to language traits, both *ROBO2* and *FOXP2* are associated with brain development and cognitive processes (Enard et al. 2002; López-Bendito et al. 2007). Previous studies have hypothesized that these genes might reside in introgression deserts, potentially indicating selective pressure to preserve specific brain functions against introgressed variants from archaic humans that could disrupt these traits (Kuhlwilm 2018; Buisan et al. 2022). However, Atkinson et al. (2018) found no evidence of recent adaptation at *FOXP2*, attributing previous signals to sample composition in which an intronic region enriched for sites that are conserved in nonhuman primates but are polymorphic among humans may be connected to a loss of function in humans. While Atkinson et al. (2018) attributed earlier findings of selection at this locus to sample composition, their methodology may have been underpowered to detect older or subtler sweeps. Flex-sweep (Lauterbur et al. 2023), a modern and sensitive method, also identifies *FOXP2* as a high-confidence candidate for positive selection. Despite the detection of sweep signals at *FOXP2* using *TrIdent*, further investigation may be warranted to reconcile these results and to better understand the evolutionary history of this gene.

To explore whether high-scoring sweep candidates were enriched for specific biological processes, molecular functions, or cellular components, we performed Gene Ontology (GO) analysis using GOrilla (Eden et al. 2009) applied to a single ranked list, where genes were ranked in decreasing order based on their highest 10-window mean sweep probabilities. Significant GO term enrichment was selected based on false discovery rate adjusted *q*-values of <0.05. The GO analysis of our CEU scan (supplementary tables S3 and S4, Supplementary Material online) identified several significant biological process terms, including regulation of cell migration, regulation of cell motility, and positive and negative regulation of locomotion, consistently involving key genes

MTOR and *LAMC2* that we highlighted earlier. Both genes are critical in pathways associated with cell movement and localization, and their dysregulation may contribute to cancer progression through mechanisms like abnormal cellular migration, localization, and growth (Kariya and Miyazaki 2004; Stipp 2010; Zoncu et al. 2011; Masuda et al. 2012; Saxton and Sabatini 2017). Additionally, the GO analysis highlighted an enrichment of genes involved in essential cellular components, especially those associated with the plasma membrane and cytoskeleton, which are central to cellular structure and movement (Alberts et al. 2002; Fletcher and Mullins 2010; Keren 2011; Jacobson et al. 2019). Alterations in these components can facilitate uncontrolled cell motility and invasion seen in metastatic cancer (Condeelis and Pollard 2006; Yilmaz and Christofori 2009). The identified cellular component GO terms also include laminin complex and cell projection part that are associated with *LAMC2*. *LAMC2* helps encode the laminin complex protein that supports cell growth, motility, and adhesion, essential processes that are often disrupted in cancer progression. In contrast, the YRI GO analysis did not yield significantly enriched terms, indicating that, unlike the CEU population, the sweep candidate genes in the YRI population do not appear to center on specific molecular functions, cellular components, or biological processes to the same extent.

Discussion

Our results show that the *TrIdent*[*IRV2*] model demonstrates strong performance in detecting selective sweeps, outperforming *T-REx*, a technique that also uses images of haplotype variation as input, in both accuracy and power. *TrIdent*[*IRV2*] effectively captured the complex genetic patterns suggestive of selective sweeps through the utilization of the pretrained *InceptionResNetV2* architecture, which is known for its sophisticated deep feature extraction prowess. Due to overall higher accuracy and power, we adopted the *Trident*[*IRV2*] model over four alternative *TrIdent* models. The complex architecture of *InceptionResNetV2* allows *TrIdent*[*IRV2*] to effectively identify signals of selection from relatively small amounts of data that may be missed by simpler models. Moreover, the use of a pretrained deep learning model as an upstream feature extractor in *TrIdent*[*IRV2*] offers practical benefits beyond just performance, including reduction in time and resources required for model training.

A key strength of *TrIdent* is its ability to achieve strong classification performance even when trained with a relatively small dataset, with the original implementation requiring only 1,000 samples per class. Notably, increasing the training data size does not yield substantial improvements in performance, suggesting that *TrIdent* effectively captures the relevant patterns with limited data. For the CEU test case, increasing the training set to 5,000 samples per class results in only a marginal 0.4% improvement in classification accuracy, with a further increase to 10,000 samples per class yielding just an additional 0.2% gain (supplementary fig. S13, Supplementary Material online). Similarly, for the YRI test case, increasing the training set to 5,000 samples per class improves accuracy by only 0.2%, and expanding it to 10,000 samples per class provides no further improvement in classification accuracy. These findings indicate that *TrIdent* reaches near-optimal performance with relatively small training datasets, making it a highly data-efficient approach for sweep detection.

The differences between the image generation strategies of *TrIdent* and *T-REx* lie in several key aspects. First, *T-REx* processes an entire 1.1 Mb region by sorting haplotypes in windows of 100 SNPs with a stride of 10 SNPs, whereas *TrIdent* focuses on a more fine-grained approach, using windows of 25 SNPs with a stride of two SNPs across a region of 499 SNPs. Second, at the SNPs where different windows overlap, *T-REx* averages the minor allele count at these SNPs after performing an L_1-norm-based haplotype rearrangement. To accomplish this task, *T-REx* first stores the windows, rearranges them, and then calculates the intersections between windows to average the minor allele counts at these intersections. This process is meant to smooth transitions between windows and retain information, especially useful for detecting weak or old sweeps. In contrast, *TrIdent* follows a more streamlined approach, which sequentially fills the input image matrix by calculating minor allele counts and sorting these counts directly within windows, avoiding the need to store windows or recalculate quantities at window intersections. This procedure not only allows *TrIdent* to bypass the additional space overhead required by *T-REx*, it also allows for faster processing times and based on the presented results, is a more computationally efficient set of operations that does not compromise the detection of selection signals. Our image generation approach therefore prevents the need for extra manipulation of genomic images, which can lead to the emergence of summary statistics and poor generalization (Cecil and Sugden 2023).

In contrast, when comparing *TrIdent* to shallow CNN-based methods that use images of summary statistic variation as input (e.g. diploS/HIC), we found that *TrIdent* trained with native images underperformed relative to these competitors. However, the nature of input data explains this disparity. In particular, diploS/HIC benefits from the usage of relatively shallow CNNs specifically trained on structured matrices of summary statistics, yet *TrIdent* outperforms diploS/HIC when trained on images of the same summary statistics. This performance boost highlights the advantage *TrIdent* gains from utilizing the *InceptionResNetV2* architecture, which excels in extracting complex patterns from images. Unlike the shallower CNNs in diploS/HIC, *InceptionResNetV2* combines both deep and wide feature extraction pathways, enabling *TrIdent* to capture genomic variation patterns in a richer and more nuanced manner.

While diploS/HIC is a widely recognized strong performer in detecting sweeps, newer and more powerful summary statistic-based tools like Flex-sweep (Lauterbur et al. 2023) have demonstrated further improvement in detection accuracy. Flex-sweep not only employs a broader set of advanced summary statistics than diploS/HIC but also incorporates custom-designed statistics that have been shown to be highly effective. A key advantage of Flex-sweep lies in its unique approach to summary statistic computation as it computes nine out of its 11 summary statistics within each window at five different nested scales. This multiscale approach allows Flex-sweep to capture genomic variation patterns more comprehensively, leading to superior performance. Despite both methods utilizing similar CNN-based architectures, the enhanced information content within the summary statistic images used by Flex-sweep likely explains its improved accuracy. Furthermore, our results with *TrIdent*[*IRV2, SS*] reinforce the idea that summary statistic-based images, when carefully designed for pretrained models, can significantly boost performance. Though *TrIdent* integrates both image generation and feature extraction-prediction methodologies, users can selectively deploy either component depending on their analytical goals, offering flexibility in genomic sweep detection tasks.

In analyzing the GradCAM-based class activation maps generated from *TrIdent*[*IRV2*], we observed a consistent focus on the lower-middle regions of the images, which corresponds to the difference of high minor allele count pattern between sweep and neutral replicates. Interestingly, a slight asymmetry in the gradient patterns was noted, with the right-hand side of the lower-center region appearing more emphasized than the left. We hypothesize that this asymmetry may arise from redundancy in the pixel patterns across the two halves of the center pattern. Specifically, the pattern on the right-hand side could render the left-hand side redundant due to the regularization penalty applied during the training of the logistic regression model. To investigate this hypothesis, we horizontally flipped the input images from the original training datasets and recomputed the GradCAM maps to generate the mean heatmaps. Despite this transformation, the resulting maps continued to emphasize the right-hand side of the lower-center pattern (supplementary fig. S14, Supplementary Material online). This consistent focus, irrespective of flipping, suggests that the model identifies key features in these regions that are central to distinguishing sweeps from neutrality, even if one-half of the central change in gradient renders the other half marginally redundant.

To evaluate the robustness of *TrIdent* in detecting sweeps across different selection strengths, we analyzed its performance across varying selection coefficient ranges. Because the original implementation of *TrIdent*[*IRV2*] was trained on sweeps with selection coefficients within the interval [0.005, 0.1], it predictably struggled to detect sweeps in the weaker selection range of [0.001, 0.005], as these values fell outside the training distribution. Specifically, sweep classification accuracy in this range was 59.15% for the CEU test case and 63.3% for the YRI test case (supplementary fig. S15, Supplementary Material online). To explore whether we can improve generalizability, we trained new *TrIdent*[*IRV2*] models on datasets where selection coefficients were drawn from the broader range of [0.001, 0.1], sampled uniformly on a logarithmic scale. This adjustment caused a slight decrease in performance on the original test dataset. Detection of sweeps in the weaker selection range of [0.001, 0.005] for the CEU test case was increased by only around 3%, yet the YRI test case improved by almost 7% (supplementary fig. S16, Supplementary Material online; right panels). While this analysis demonstrates that the performance of *TrIdent* in detecting weaker sweeps is contingent upon the selection regime used during training, we hypothesize that we may need to substantially increase training samples to further improve generalizability of *TrIdent* in regard to detecting both stronger and weaker sweeps.

Low-coverage data presents significant challenges for genomic analysis due to the uncertainty in determining genotypes (Nielsen et al. 2011; Fumagalli 2013; Korneliussen et al. 2014). This uncertainty stems from various sources, including mapping errors, sequencing errors, and the random sampling of haploid reads from a diploid genotype, making it difficult to accurately infer the underlying genetic variation. In genomic regions of low coverage, the detection of selective sweeps is particularly problematic, as the likelihood of missing key genetic variants increases, leading to incomplete or inaccurate

genomic signals that may masquerade as sweeps. *TrIdent* may be well-suited for applications to low-coverage data, as it relies on windows of minor allele counts and works with unphased multilocus genotype data. These minor allele counts could be approximated with expected allele counts from genotype likelihoods (Fumagalli et al. 2014; Fox et al. 2019), which would permit *TrIdent* to account for the uncertainties in calling genotypes from low-coverage sequencing and has been shown to be a suitable approach for similar settings (Gower et al. 2021). Finally, using minor allele counts within *TrIdent* input images is advantageous as it does not require a reliable outgroup sequence to polarize alleles as derived or ancestral. This characteristic is important, as erroneous allele polarization can alter the distribution of allele frequencies to be skewed toward high-frequency derived alleles, much like that of a selective sweep (Hernandez et al. 2007).

To expand upon how our local sorting of haplotypes influences model performance, we explored its benefits compared to a global sorting approach, which we denote as *TrIdent*[IRV2, *global*]. The image generation process of *TrIdent*, as described in the *Image Generation* subsection of the *Materials and Methods*, arranges haplotypes based on minor allele counts independently within small sub-windows, preserving localized structural patterns critical for classification. In contrast, global sorting organizes the entire 499-SNP window at once based on total minor allele counts. While global sorting guarantees that every pixel value in a row of the resulting image corresponds to the same haplotype, this uniform arrangement may obscure subtle, localized patterns essential for distinguishing sweeps from neutrality. Our results demonstrate that local sorting significantly improves predictive accuracy, with global sorting resulting in accuracy losses of 9.25% in the CEU dataset and 6.1% in the YRI dataset (supplementary fig. S17, Supplementary Material online). The larger effect in the CEU population, which has experienced a bottlenecked demographic history, highlights the critical role of local patterns that are more susceptible to being blurred by global sorting. By preserving the fine-grained haplotype structures within sub-windows, local sorting enables *TrIdent* to capture nuanced signals associated with selective sweeps, especially in populations with complex evolutionary histories.

In our efforts to bridge the gap between the highly complex deep CNN architecture of *InceptionResNetV2* and comparatively shallower sequential CNN architecture used in *smbCNN* (which is identical to the diploS/HIC architecture), we examined a custom designed residual (He et al. 2016) and multipath (Szegedy et al. 2017) CNN architecture that we term *scCNN* (shallow complex CNN) for ease of reference. By incorporating advanced architectural features, *scCNN* enhances complex feature extraction and maintains training efficiency, making it well-suited for complex tasks without the need for excessively deep networks like *InceptionResNetV2*. The architecture of *scCNN* is detailed in *Constructing the scCNN architecture* in *Methods*. Despite the structural complexities introduced in *scCNN* compared to *smbCNN*, *scCNN* achieved accuracies of 84.10% on the CEU dataset and 86.75% on the YRI dataset (supplementary fig. S18, Supplementary Material online). The *scCNN* model still falls short of the performance levels achieved by *TrIdent*[IRV2] with its pretrained *InceptionResNetV2* backbone. The reason lies in the inherent strengths of the *InceptionResNetV2* architecture, which benefits from extensive pretraining on large-scale

datasets and a more sophisticated multipath design that can capture a broader range of features across different scales. While this *scCNN* model represents a marginal improvement over the simpler *smbCNN* architecture, it underscores the superior capability of the *InceptionResNetV2* architecture employed by *TrIdent*[IRV2] in detecting subtle patterns of selection in our haplotype rearrangement-based images.

Though *TrIdent* achieved good performance on demographic settings with fluctuating populations, it is also important to consider other, more complex, demographic factors that could impact method performance, such as significant population substructure and admixture. These factors, particularly the timing and scale of migration events, can strongly influence the observed spatial haplotypic diversity and may result in misleading indicators of selection (Harris et al. 2018). The mean of the sweep images used by *TrIdent* (supplementary fig. S1, Supplementary Material online) reveals a funnel-shaped dark region at the selected locus (center columns of the image), indicating low pixel intensity values that are absent in the mean of the neutral images. Class activation maps (Fig. 7) also show that *TrIdent* focuses on the lower-middle input image pixels for classification. Now, the image generation process (Fig. 1) illustrates that pixels at the top of the images represent haplotypes with low minor allele counts resulting in zero or near-zero values, whereas pixels toward the bottom of the images reflect haplotypes with higher minor allele counts resulting in higher values that can reach values up to 25 (*Modeling Description*). This observation suggests that the loss of haplotype diversity near sweep loci leads to more near-zero values in the middle-to-lower parts of the image, which *TrIdent* uses for classification. Therefore, any form of migration or population structure introducing near-zero minor allele counts in this region could be misclassified as a sweep, whereas higher values could obscure sweep signals. Migration from a population with a smaller effective size can introduce haplotypes with limited diversity to the recipient population. As a consequence, high migration rate settings may replace much of the haplotypic variation in the recipient with this low haplotype diversity from the donor, leading many polymorphic sites with low minor allele counts, thereby increasing the occurrence of near-zero values in the middle-to-lower sections of image matrices and potentially causing false inferences of sweeps. On the other hand, such migration from a population with a moderate to large effective size after the beneficial allele has swept can mask sweep footprints by replacing the low haplotypic diversity with moderate to high levels from the donor population, resulting in higher pixel values in the middle-to-lower image regions. Therefore, if it is expected that the study population may have received significant gene flow from other populations, then accounting for such processes when generating training datasets is likely important to guard against both false positive and false negative results when detecting sweeps.

Furthermore, we conducted additional analyses to assess the impact of sweep shoulders on the classification performance of *TrIdent*. Specifically, we evaluated the ability of *TrIdent* to detect sweeps when the beneficial mutation was positioned at different locations within a 1.1 Mb sequence. By centering sweeps at 500, 375, 250, 125, and 0 kb away from the center, we found that *TrIdent* gradually loses detection ability when the sweep moves farther off-center for both CEU and YRI test cases (supplementary fig. S19, Supplementary Material online; left panel). For example, when the sweep is positioned 125 kb

away from the image center, sweep detection accuracy is approximately 78%, but performance predictably drops substantially at greater distances. To further investigate how sweep probability signals diffuse across the genomic region, we conducted a scan across the simulated sweep sequences, protocol of which is detailed in the *Assessing sweep shoulders* subsection of the *Methods*. This analysis enabled us to systematically assess how predicted sweep probabilities change with distance from the beneficial mutation (supplementary fig. S19, Supplementary Material online; right panel), demonstrating that while *TrIdent* can detect selective sweeps beyond a perfectly centered mutation, its accuracy declines as the focal mutation shifts further from the image center, emphasizing the importance of generating images in small genomic distances when analyzing empirical data.

The pretrained *InceptionResNetV2* model is especially suitable for assessing the image of multilocus variation used by *TrIdent* because of its hybrid architecture, which effectively integrates the advantageous qualities of both Inception and Residual networks. Inception modules facilitate the extraction of features at several scales, enabling the detection of patterns of different sizes and complexities in the input images. The added residual connections enhance the feature extraction process by utilizing network depth, as model layers have the ability to capture diverse sets of features, which allows for accurate identification of small changes in genomic variation with multilevel inspection and concatenation of information from input images without the need for further training. The promising performance of *TrIdent* on detecting sweeps when confronted with common empirical hurdles while also having the capacity to predict evolutionary parameters underlying sweeps makes it a versatile tool for population genetic studies. This versatility is underscored by consistently good performance of *TrIdent* across two demographic histories, showing marginally higher accuracy in the YRI dataset due to greater genetic diversity, whereas on the CEU dataset, affected by a recent severe bottleneck, exhibited reduced background variation that slightly diminished sweep detection accuracy. The robust performance of *TrIdent* when faced with nonsweep patterns that can masquerade as sweeps, such as technical issues that lead to large missing genomic segments and the process of background selection further underscores its ability to handle challenging genomic scenarios effectively. Missing data, such as undiscovered polymorphism due to poor mappability, can mimic selective sweeps by reducing local haplotypic diversity. Despite this challenge, *TrIdent* maintains high and only marginally-decreased accuracy under such settings. Additionally, *TrIdent* shows resilience against misleading patterns of lost diversity due to background selection. By analyzing haplotype distributions, the model can effectively guard against false classification of background selection as a selective sweep while maintaining a low FPR.

Though our empirical filtration method, as detailed in the *Filtration of empirical data and empirical image generation* subsection of the *Materials and Methods*, is effective in reducing false discoveries, it can also inadvertently exclude some true signals from the analysis. A notable case study of this phenomenon is the *ACKR1* gene (Atypical Chemokine Receptor 1), also known as the *DARC* gene (Duffy Antigen Receptor for Chemokines), which has been widely recognized as a target of positive selection due to its role in conferring resistance to malaria caused by *Plasmodium vivax* infection (Horuk 2015; Yin et al. 2018). This highly recombining locus poses a particular challenge for detecting selection signals due to its complex genetic landscape (McManus et al. 2017; Lauterbur et al. 2023). In our analysis, *ACKR1* was excluded during the filtration process because it exhibited a low mean CRG score. To assess whether this exclusion impacted the detection of a true signal at this locus, we reanalyzed SNPs within and around *ACKR1*. This reanalysis revealed evidence of a sweep in the YRI scan, with a mean peak probability of 0.858 (supplementary fig. S20, Supplementary Material online). However, no similar signal was detected in the CEU scan. This example illustrates the tradeoffs inherent in applying stringent filtration criteria and highlights the importance of carefully balancing false discovery control with the retention of biologically meaningful signals.

Finally, we scanned two human populations (CEU and YRI) for sweeps using *TrIdent*, and reported candidate genes with previous literature support, as well as some novel candidate loci. However, it is also important to highlight those candidates that *TrIdent* identified with highest confidence. We therefore considered the 10 sweep candidates (supplementary tables S1 and S2, Supplementary Material online) with highest 10-window mean sweep probabilities within each population to better compare and contrast the genomic regions strongly favored by *TrIdent*. Two genes (*LAMC2* and *MTOR*) in our CEU analysis and two genes (*HLA-DRB6* and *ROBO2*) in our YRI analysis are among the top 10 candidates in their respective scans. By focusing on these top 10 candidates, which have no overlap between the pair of sweep scans, we may gain insights into the key genetic adaptations and distinct evolutionary pressures that have shaped these populations. Whereas the CEU candidates reflect adaptations potentially linked to metabolism, neurodevelopment, and cancer-related pathways, those from the YRI focus more on immune function and pathogen resistance. This potential functional divergence highlights how different environments and historical pressures may have shaped the genetic landscape of these populations in unique ways. However, the exclusivity of candidates may also suggest a limitation of the trained *TrIdent* model, in which it may have reduced capacity for detecting more ancient sweeps that would have occurred prior to the split of CEU and YRI, thus missing shared sweep signals across these populations. Future studies could train *TrIdent* models on variation across multiple populations to detect shared sweeps and discriminate them from population-specific events.

Materials and Methods

Simulation Protocol

To simulate replicates under both the European and sub-Saharan African human demographic histories, we drew the per-site per-generation mutation rate μ uniformly at random within the interval $[2.21 \times 10^{-9}, 2.21 \times 10^{-8}]$ with a mean of 1.21×10^{-8} (Scally and Durbin 2012; Schrider and Kern 2017) and the per-site per-generation recombination rate r at random from an exponential distribution with mean 10^{-8} (Payseur and Nachman 2000; Schrider and Kern 2017) and truncated at three times the mean (Schrider and Kern 2017). For each replicate simulated under the inferred demographic history (Tennessen et al. 2012), we sampled 198 haplotypes for CEU and 216 haplotypes for YRI of length 1.1 Mb to match the number of sampled haplotypes in our empirical experiments.

We generated sweep simulations with per-generation selection strength s acting on a beneficial allele at frequency f that arose at a locus at the center of a simulated genomic region, with s drawn uniformly at random within the interval $[0.005, 0.1]$ and f drawn uniformly at random on a logarithmic scale within the interval $[1/(2N_e), 0.2]$ (Schrider and Kern 2017), where N_e is the present-day effective population size. Moreover, we also drew the number of generations τ in the past at which the beneficial allele reached fixation uniformly at random within the interval $[0, 2,000]$ (Schrider and Kern 2017). This simulation protocol guarantees that both training and test sets reflect an array of sweep settings, including weaker (small s) and stronger (large s) selection scenarios, harder (small f) and softer (large f) sweeps, and time τ in which mutation, recombination, and genetic drift can erode signals of sweeps, all of which affect the size and prominence of the genomic footprint left by a sweep after fixation. These modeled genetic, demographic, and selection parameters are expected to lead to significant overlap in the distributions of genetic variation between sweeps and neutrality.

Image Generation

For each simulated replicate, we retain only bi-allelic SNPs with a minor allele count of at least three (i.e. removed singletons and doubletons). We then generate a matrix $\mathbf{M}$ of dimension $n \times 499$, where n denotes the number of haplotypes in a replicate, with each row representing one of the sampled haplotypes and each column representing one of 499 SNPs, where these SNPs were chosen as the central position, 249 closest SNPs upstream of the central position, and 249 closest SNPs downstream of the central position of the simulated genomic region, which is where the selected mutation is introduced in sweep simulations. The element $\mathbf{M}_{ij}$ takes a value of zero if the haplotype within the ith row has the major allele at the SNP in the jth column, and otherwise has the value of one for the minor allele. From this matrix, we sought to create a representation of genome variation that might make pattern recognition easier by generating a matrix $\mathbf{X}$ of dimension $n \times 237$ computed using windows of 25 SNPs with a stride of two between each window, with element values $\mathbf{X}_{ij} \in \{0, 1, \ldots, 25\}$. That is, for $j \in \{1, 2, \ldots, 237\}$, we computed the minor allele counts for each of the n rows of a genomic window of length 25 SNPs starting at column $2j - 1$ of $\mathbf{M}$, arranged these minor allele counts values in increasing order, and set the values in column j of $\mathbf{X}$ as this sorted list of minor allele counts. This procedure summarizes the number of minor alleles within a given 25-SNP window for each haplotype, and thus rows toward the top of $\mathbf{X}$ summarize haplotypes with a greater number of major alleles, whereas rows toward the bottom have a greater number of minor alleles. We then perform the same computation of each 25-SNP window by taking a stride of two SNPs. This matrix is subsequently resized using linear interpolation, depending on the input image size requirements of the pretrained neural network architectures. For example, pretrained models like *InceptionResNetV2* require input image sizes of 299×299, whereas for some pretrained models, such as *VGG16* or *MobileNetV2*, the required input image size is 224×224. An example illustrating this image generation procedure is presented in Fig. 1 (top panel).

Constructing the *scCNN* Architecture

The *scCNN* architecture begins with an initial convolutional layer that applies a set of filters to the input image.

Following this layer, the model incorporates a residual block that contains two convolutional layers, each followed by batch normalization (Bjorck et al. 2018) and ReLU activation (Krizhevsky 2012). A skip connection within the residual block bypasses these two convolution layers, directly adding the input to the next convolution layer, which maintains gradient flow during training and enables deeper feature learning (He et al. 2016). There is an additional skip connection before the fifth convolution layer that allows the concatenation of fourth and fifth convolution layer outputs that are then fed into the dense layer. The model is optimized using the Adam optimizer and binary cross-entropy loss, ensuring efficient training and robust performance.

Computational Resources and Requirements

Our analysis was conducted on a system equipped with an AMD EPYC 7,702 64-core CPU with 100 GB of RAM. The *TrIdent* image generation method requires a mean of 6.83 s to generate an image from a simulated replicate. Loading the *InceptionResNetV2* architecture with its pretrained weights uses approximately 513 MB of RAM. The mean time to compute the GAP layer output using *InceptionResNetV2* is 0.018 s for a single input image. Training the penalized logistic regression model on 2,000 images with hyperparameter tuning, if run parallelly, requires approximately 6 min and 30 s and consumes 56 MB of memory.

Conversely, training the entire *InceptionResNetV2* architecture, as executed in the training of the *IRV2* model, necessitates additional resources beyond those previously mentioned. Specifically, training *IRV2* on 2,000 images, incorporating early stopping to prevent overfitting, attained convergence in 37 epochs for the CEU dataset and 26 epochs for the YRI dataset. Each epoch required approximately 4 min to complete and utilized around 26 GB of RAM in total for both test cases. We anticipate that as more training images are used, the RAM utilization and the training time would increase substantially for the full *IRV2* model.

For comparison, we also assessed the computational resources required by diploS/HIC. Extracting 101 summary statistics from each simulated replicate takes on average 17.17 s. Training the CNN model with early stopping, using a dataset consisting of summary statistics from 1,000 replicates per class, allows diploS/HIC to converge in 13 epochs, requiring approximately 42 min and consuming 7.5 GB of RAM.

Application of CLUES2

To benchmark the accuracy and precision of the *TrIdent*[*IRV2*, *ANN*] nonlinear regression model in estimating selection coefficients, we compared its results against CLUES2 (Vaughn and Nielsen 2024). First, we used the `ms2vcf` tool of the `coatli` (Klassmann 2013) package to convert `discoal` output ms formatted files to VCF format, as SINGER (Deng et al. 2024), one of the two software supported by CLUES2 to compute ARGs, accepts VCF files as input. Because the exact mutation rate applied in each simulated replicate is unavailable, we used a mean mutation rate of 1.21×10^{-8}, consistent with our simulations. Additionally, as we could not extract the recombination to mutation rate ratio for each replicate, we set this parameter to the default value of one in SINGER. We also retained the default number of posterior samples of 100. After generating ARGs, we used the `SingerToCLUES` pipeline of CLUES2 to convert the

output from SINGER to the required CLUES2 input format. For the genomic position of the beneficial mutation, we selected the SNP closest to the simulation center. Finally, to run the CLUES2 inference script, we used the sample-derived allele frequency of this central SNP. For the maximum time (in generations) to be considered in the analysis, we chose 2,000 generations in the past. We also tested 1,000 and 4,000 generations, but the results remained largely unchanged.

Assessing Sweep Shoulders

To analyze classification performance of *TrIdent* on sweep shoulders, for both CEU and YRI datasets, we computed the mean sweep probability across 1,000 test images within overlapping 20 kb bins spanning the 1.1 Mb simulated region. Using a stride of 10 kb, we created 48 bins with the first bin spanning 300 to 320 kb to last bin spanning 780 to 800 kb. We exclude the first and last 300 kb of the region because the image generation process requires at least 499 SNPs with the center SNP falling within a bin. When a bin is positioned too far from the simulation center (550 kb), the number of qualifying images meeting these criteria becomes insufficient for reliable analysis. To generate images for this experiment, we scanned the entire 1.1 Mb sweep replicates in an identical manner to how images were generated from chromosomes in our empirical analysis, as described in the *Filtration of empirical data and empirical image generation* subsection. Once generated, images were assigned to bins based on their center SNP positions. Any image with a center SNP falling outside these predefined bins was excluded from further analysis.

Filtration of Empirical Data and Empirical Image Generation

To apply *TrIdent* to the phased genotype data from the 99 CEU and 108 YRI individuals of the 1000 Genomes Project dataset (1000 Genomes Project Consortium 2015), we retained only bi-allelic SNPs with a minor allele count of at least three (i.e. removed singletons and doubletons). Following the protocol of Mughal et al. (2020), we further filtered this dataset by excluding genomic segments of length 100 kb with mean CRG (Consensus Reference Genomes) mappability and alignability scores (Talkowski et al. 2011) below 0.9 to reduce the possibility of misleading signals due to technical concerns. CRG mappability scores reflect the probability of accurately mapping short sequencing reads to a specific genomic region. We then created images for input to *TrIdent* by considering the first 499 contiguous SNPs on an autosome, converting the haplotype variation across these SNPs into an image according to the procedure outlined in the *Image generation* subsection, processing subsequent images by moving the SNP window by a stride of two SNPs along the autosome, and repeating this procedure for all autosomes. The chromosomal location of each observation was set as mean position of the 249th and 250th SNP.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

Computations for this research were performed using the services provided by Research Computing at the Florida Atlantic University.

Data Availability

We release the source code for *TrIdent* under the MIT open source license, and this repository can be accessed on GitHub (https://github.com/sandipanpaul06/TrIdent/). To facilitate reuse and encourage further analysis, we have released the weights and architecture of the *IRV2* models trained on CEU and YRI simulations through Zenodo (https://doi.org/10.5281/zenodo.15225972). The CEU and YRI data from the 1000 Genomes Project can be accessed from the project website (https://www.internationalgenome.org/category/phase-3/).

References

1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015:526(7571):68. https://doi.org/10.1038/nature15393.

Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado G, Davis A, Dean J, Devin M, *et al.* 2015. TensorFlow: large-scale machine learning on heterogeneous systems. Software available from tensorflow.org. https://www.tensorflow.org/.

Adrion J, Cole C, Dukler N, Galloway J, Gladstein A, Gower G, Kyriazis C, Ragsdale A, Tsambos G, Baumdicker G, *et al.* A community-maintained standard library of population genetic models. *eLife*. 2020:9:e54967. https://doi.org/10.7554/eLife.54967.

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. The cytoskeleton and cell behavior. In: *Molecular biology of the cell*. 4th ed. New York (NY): Garland Science; 2002.

Albrechtsen A, Moltke I, Nielsen R. Natural selection and the distribution of identity-by-descent in the human genome. *Genetics*. 2010:186(1):295–308. https://doi.org/10.1534/genetics.110.113977.

Alkan C, Sajjadian S, Eichler E. Limitations of next-generation genome sequence assembly. *Nat Methods*. 2011:8(1):61–65. https://doi.org/10.1038/nmeth.1527.

Allison A. Protection afforded by sickle-cell trait against subtertian malarial infection. *Br Med J*. 1954:1(4857):290. https://doi.org/10.1136/bmj.1.4857.290.

Alzubaidi L, Zhang J, Humaidi A, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel M, Al-Amidie M, Farhan L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data*. 2021:8(1):1–74. https://doi.org/10.1186/s40537-021-00444-8.

Amin M, Hasan M, Arnab S, DeGiorgio M. Tensor decomposition-based feature extraction and classification to detect natural selection from genomic data. *Mol Biol Evol*. 2023:40(10):msad216. https://doi.org/10.1093/molbev/msad216.

Anguita-Ruiz A, Aguilera C, Gil A. Genetics of lactose intolerance: an updated review and online interactive world maps of phenotype and genotype frequencies. *Nutrients*. 2020:12(9):2689. https://doi.org/10.3390/nu12092689.

Arnab S, Amin M, DeGiorgio M. Uncovering footprints of natural selection through spectral analysis of genomic summary statistics. *Mol Biol Evol*. 2023:40(7):msad157. https://doi.org/10.1093/molbev/msad157.

Atkinson E, Audesse A, Palacios J, Bobo D, Webb A, Ramachandran S, Henn B. No evidence for recent selection at FOXP2 among diverse human populations. *Cell*. 2018:174(6):1424–1435. https://doi.org/10.1016/j.cell.2018.06.048.

Azizpour H, Razavian A, Sullivan J, Maki A, Carlsson S. Factors of transferability for a generic ConvNet representation. *IEEE Trans Pattern Anal Mach*. 2015:38(9):1790–1802. https://doi.org/10.1109/TPAMI.2015.2500224.

Barton N, Charlesworth B. Why sex and recombination? *Science*. 1998:281(5385):1986–1990. https://doi.org/10.1126/science.281.5385.1986.

Battey C, Ralph P, Kern A. Predicting geographic location from genetic variation with deep neural networks. *eLife*. 2020:9:e54507. https://doi.org/10.7554/eLife.54507.

Bersaglieri T, Sabeti P, Patterson N, Vanderploeg T, Schaffner S, Drake JA, Rhodes M, Reich D, Hirschhorn J. Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet*. 2004:74(6):1111–1120. https://doi.org/10.1086/421051.

Beyer L, Hénaff O, Kolesnikov A, Zhai X, Oord A. 2020. Are we done with imagenet? arXiv, arXiv:2006.07159, preprint: not peer reviewed.

Bjorck J, Gomes C, Selman B, Weinberger KQ. Understanding batch normalization. Adv Neural Inf Process Syst. 2018:31.

Boyko A, Williamson S, Indap A, Degenhardt J, Hernandez R, Lohmueller K, Adams M, Schmidt S, Sninsky J, Sunyaev S, *et al*. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet*. 2008:4(5):e1000083. https://doi.org/10.1371/journal.pgen.1000083.

Bozinovski S. Reminder of the first paper on transfer learning in neural networks, 1976. *Informatica*. 2020:44(3):291–302. https://doi.org/10.31449/inf.v44i3.2828.

Buisan R, Moriano J, Andirkó A, Boeckx C. A brain region-specific expression profile for genes within large introgression deserts and under positive selection in homo sapiens. *Front Cell Dev Biol*. 2022:10:824740. https://doi.org/10.3389/fcell.2022.824740.

Cecil R, Sugden L. On convolutional neural networks for selection inference: revealing the effect of preprocessing on model learning and the capacity to discover novel patterns. *PLoS Comput Biol*. 2023:19(11):e1010979. https://doi.org/10.1371/journal.pcbi.1010979.

Cesaratto L, Grisard E, Coan M, Zandonà L, De Mattia E, Poletto E, Cecchin E, Puglisi F, Canzonieri V, Mucignat MT, *et al*. BNC2 is a putative tumor suppressor gene in high-grade serous ovarian carcinoma and impacts cell survival after oxidative stress. *Cell Death Dis*. 2016:7(9):e2374–e2374. https://doi.org/10.1038/cddis.2016.278.

Chan J, Perrone V, Spence J, Jenkins P, Mathieson S, Song Y. A likelihood-free inference framework for population genetic data using exchangeable neural networks. *Adv Neural Inf Process Syst*. 2018:31:8594–8605. https://arxiv.org/abs/1802.06153.

Charlesworth B. The effects of deleterious mutations on evolution at linked sites. *Genetics*. 2012:190(1):5–22. https://doi.org/10.1534/genetics.111.134288.

Charlesworth B, Morgan M, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. *Genetics*. 1993:134(4):1289–1303. https://doi.org/10.1093/genetics/134.4.1289.

Charlesworth B, Nordborg M, Charlesworth D. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet Res*. 1997:70(2):155–174. https://doi.org/10.1017/S0016672397002954.

Charlesworth D, Charlesworth B, Morgan M. The pattern of neutral molecular variation under the background selection model. *Genetics*. 1995:141(4):1619–1632. https://doi.org/10.1093/genetics/141.4.1619.

Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. *Genome Res*. 2010:20(3):393–402. https://doi.org/10.1101/gr.100545.109.

Cheng X, Xu C, DeGiorgio M. Fast and robust detection of ancestral selective sweeps. *Mol Ecol*. 2017:26(24):6871–6891. https://doi.org/10.1111/mec.2017.26.issue-24.

Comeron J. Background selection as baseline for nucleotide variation across the drosophila genome. *PLoS Genet*. 2014:10(6):e1004434. https://doi.org/10.1371/journal.pgen.1004434.

Condeelis J, Pollard J. Macrophages: obligate partners for tumor cell migration, invasion, and metastasis. *Cell*. 2006:124(2):263–266. https://doi.org/10.1016/j.cell.2006.01.007.

Costa C, Pereira S, Lima L, Peixoto A, Fernandes E, Neves D, Neves M, Gaiteiro C, Tavares A, Gil da Costa R, *et al*. Abnormal protein glycosylation and activated PI3K/Akt/mTOR pathway: role in bladder cancer prognosis and targeted therapeutics. *PLoS One*. 2015:10(11):e0141253. https://doi.org/10.1371/journal.pone.0141253.

DeGiorgio M, Szpiech Z. A spatially aware likelihood test to detect sweeps from haplotype distributions. *PLoS Genet*. 2022:18(4):e1010134. https://doi.org/10.1371/journal.pgen.1010134.

Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: *CVPR09*. Los Alamitos (CA): IEEE Computer Society; 2009.

Deng Y, Nielsen R, Song Y. Robust and accurate Bayesian inference of genome-wide genealogies for large samples. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society. p. 248–255.

Di Rienzo A, Hudson R. An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends Genet*. 2005:21(11):596–601. https://doi.org/10.1016/j.tig.2005.08.007.

Di Rienzo S. Population genetics models of common diseases. *Curr Opin Genet Dev*. 2006:16(6):630–636. https://doi.org/10.1016/j.gde.2006.10.002.

Duffy DL, Montgomery GW, Chen W, Zhao ZZ, Le L, James MR, Hayward NK, Martin NG, Sturm RA. A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. *Am J Hum Genet*. 2007:80(2):241–252. https://doi.org/10.1086/510885.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. Gorilla: a tool for discovery and visualization of enriched go terms in ranked gene lists. *BMC Bioinformatics*. 2009:10(1):1–7. https://doi.org/10.1186/1471-2105-10-48.

Enard W, Przeworski M, Fisher S, Lai C, Wiebe V, Kitano T, Monaco A, Pääbo S. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*. 2002:418(6900):869–872. https://doi.org/10.1038/nature01025.

Fagny M, Patin E, Enard D, Barreiro L, Quintana-Murci L, Laval G. Exploring the occurrence of classic selective sweeps in humans using whole-genome sequencing data sets. *Mol Biol Evol*. 2014:31(7):1850–1868. https://doi.org/10.1093/molbev/msu118.

Fay J, Wu C. Sequence divergence, functional constraint, and selection in protein evolution. *Annu Rev Genomics Hum Genet*. 2003:4(1):213–235. https://doi.org/10.1146/genom.2003.4.issue-1.

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol Biol Evol*. 2014:31(5):1275–1291. https://doi.org/10.1093/molbev/msu077.

Field Y, Boyle E, Telis N, Gao Z, Gaulton K, Golan D, Yengo L, Rocheleau G, Froguel P, McCarthy M, *et al*. Detection of human adaptation during the past 2000 years. *Science*. 2016:354(6313):760–764. https://doi.org/10.1126/science.aag0776.

Fijarczyk A, Babik W. Detecting balancing selection in genomes: limits and prospects. *Mol Ecol*. 2015:24(14):3529–3545. https://doi.org/10.1111/mec.2015.24.issue-14.

Flagel L, Brandvain Y, Schrider DR. The unreasonable effectiveness of convolutional neural networks in population genetic inference. *Mol Biol Evol*. 2019:36(2):220–238. https://doi.org/10.1093/molbev/msy224.

Fletcher D, Mullins R. Cell mechanics and the cytoskeleton. *Nature*. 2010:463(7280):485–492. https://doi.org/10.1038/nature08908.

Fox E, Wright A, Fumagalli M, Vieira F. ngsLD: evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics*. 2019:35(19):3855–3856. https://doi.org/10.1093/bioinformatics/btz200.

Fumagalli M. Assessing the effect of sequencing depth and sample size in population genetics inferences. *PLoS One*. 2013:8(11):e79667. https://doi.org/10.1371/journal.pone.0079667.

Fumagalli M, Vieira F, Linderoth T, Nielsen R. ngsTools: methods for population genetics analyses from next-generation sequencing

data. *Bioinformatics*. 2014:30(10):1486–1487. https://doi.org/10.1093/bioinformatics/btu041.

Garud N, Messer P, Buzbas E, Petrov D. Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. *PLoS Genet*. 2015:11(2):e1005004. https://doi.org/10.1371/journal.pgen.1005004.

Gerbault P, Moret C, Currat M, Sanchez-Mazas A. Impact of selection and demography on the diffusion of lactase persistence. *PLoS One*. 2009:4(7):e6369. https://doi.org/10.1371/journal.pone.0006369.

Goeury T, Creary L, Brunet L, Galan M, Pasquier M, Kervaire B, Langaney A, Tiercy J, Fernández-Viña M, Nunes J, *et al*. Deciphering the fine nucleotide diversity of full HLA class I and class II genes in a well-documented population from sub-Saharan Africa. *HLA*. 2018:91(1):36–51. https://doi.org/10.1111/tan.2018.91.issue-1.

Gower G, Picazo P, Fumagalli M, Racimo F. Detecting adaptive introgression in human evolution using convolutional neural networks. *eLife*. 2021:10:e64669. https://doi.org/10.7554/eLife.64669.

Grant P, Grant B. Unpredictable evolution in a 30-year study of Darwin's finches. *Science*. 2002:296(5568):707–711. https://doi.org/10.1126/science.1070315.

Grealey J, Lannelongue L, Saw W, Marten J, Méric G, Ruiz-Carmona S, Inouye M. The carbon footprint of bioinformatics. *Mol Biol Evol*. 2022:39(3):msac034. https://doi.org/10.1093/molbev/msac034.

Haller B, Messer P. SLiM 3: forward genetic simulations beyond the Wright–Fisher model. *Mol Biol Evol*. 2019:36(3):632–637. https://doi.org/10.1093/molbev/msy228.

Harris A, DeGiorgio M. A likelihood approach for uncovering selective sweep signatures from haplotype data. *Mol Biol Evol*. 2020:37(10):3023–3046. https://doi.org/10.1093/molbev/msaa115.

Harris A, Garud N, DeGiorgio M. Detection and classification of hard and soft sweeps from unphased genotypes by multilocus genotype identity. *Genetics*. 2018:210(4):1429–1452. https://doi.org/10.1534/genetics.118.301502.

Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*. 2nd ed. New York (NY): Springer; 2009.

He K, Zhang X, Ren S, Sun J. 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos (CA): IEEE Computer Society. p. 770–778.

Hedrick P, Thomson G. Evidence for balancing selection at HLA. *Genetics*. 1983:104(3):449–456. https://doi.org/10.1093/genetics/104.3.449.

Hendrycks D, Lee K, Mazeika M. 2019. Using pre-training can improve model robustness and uncertainty. In: International Conference on Machine Learning. Brookline (MA): Proceedings of Machine Learning Research (PMLR). p. 2712–2721.

Hermisson J, Pennings P. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*. 2005:169(4):2335–2352. https://doi.org/10.1534/genetics.104.036947.

Hernandez R, Williamson S, Bustamante C. Context dependence, ancestral misidentification, and spurious signatures of natural selection. *Mol Biol Evol*. 2007:24(8):1792–1800. https://doi.org/10.1093/molbev/msm108.

Hoffmann A, Sgrò C. Climate change and evolutionary adaptation. *Nature*. 2011:470(7335):479–485. https://doi.org/10.1038/nature09670.

Hornik K. Approximation capabilities of multilayer feedforward networks. *Neural Netw*. 1991:4(2):251–257. https://doi.org/10.1016/0893-6080(91)90009-T.

Horuk R. The Duffy antigen receptor for chemokines DARC/ACKR1. *Front Immunol*. 2015:6:279. https://doi.org/10.3389/fimmu.2015.00279.

Huber C, DeGiorgio M, Hellmann I, Nielsen R. Detecting recent selective sweeps while controlling for mutation rate and background selection. *Mol Ecol*. 2016:25(1):142–156. https://doi.org/10.1111/mec.2016.25.issue-1.

Hudson R, Kaplan N. Deleterious background selection with recombination. *Genetics*. 1995:141(4):1605–1617. https://doi.org/10.1093/genetics/141.4.1605.

Jacobs L, Wollstein A, Lao O, Hofman A, Klaver C, Uitterlinden A, Nijsten T, Kayser M, Liu F. Comprehensive candidate gene study highlights UGT1A and BNC2 as new genes determining continuous skin color variation in Europeans. *Hum Genet*. 2013:132(2):147–158. https://doi.org/10.1007/s00439-012-1232-9.

Jacobson K, Liu P, Lagerholm B. The lateral organization and mobility of plasma membrane components. *Cell*. 2019:177(4):806–819. https://doi.org/10.1016/j.cell.2019.04.018.

Jiang J, Yu H, Shou Y, Neale G, Zhou S, Lu T, Sorrentino B. Hemgn is a direct transcriptional target of HOXB4 and induces expansion of murine myeloid progenitor cells. *Blood J Am Soc Hematol*. 2010:116:711–719. https://doi.org/10.1182/blood-2009-07-235341.

Kariya Y, Miyazaki K. The basement membrane protein laminin-5 acts as a soluble cell motility factor. *Exp Cell Res*. 2004:297(2):508–520. https://doi.org/10.1016/j.yexcr.2004.03.044.

Keinan A, Reich D. Human population differentiation is strongly correlated with local recombination rate. *PLoS Genet*. 2010:6(3):e1000886. https://doi.org/10.1371/journal.pgen.1000886.

Kelly J. A test of neutrality based on interlocus associations. *Genetics*. 1997:146(3):1197–1206. https://doi.org/10.1093/genetics/146.3.1197.

Keren K. Cell motility: the integrating role of the plasma membrane. *Eur Biophys J*. 2011:40(9):1013–1027. https://doi.org/10.1007/s00249-011-0741-0.

Kern A, Schrider D. Discoal: flexible coalescent simulations with selection. *Bioinformatics*. 2016:24(24):3839–3841. https://doi.org/10.1093/bioinformatics/btw556.

Kern A, Schrider D. diploS/HIC: an updated approach to classifying selective sweeps. *G3 (Bethesda)*. 2018:8(6):1959–1970. https://doi.org/10.1534/g3.118.200262.

Kim Y, Nielsen R. Linkage disequilibrium as a signature of selective sweeps. *Genetics*. 2004:167(3):1513–1524. https://doi.org/10.1534/genetics.103.025387.

Kimura M. The neutral theory of molecular evolution. *Sci Am*. 1979:241(5):98–129. https://doi.org/10.1038/scientificamerican1179-98.

Kingma D, Ba J. 2015. Adam: a method for stochastic optimization 3rd International Conference on Learning Representations. In: ICLR 2015-Conference Track Proceedings. San Diego (CA): International Conference on Learning Representations (ICLR). Vol. 1.

Klassmann A. 2013. Coatli: a software package for computational analysis. Software available under GNU General Public License version 3.0 (GPLv3). http://sourceforge.net/p/coatli/wiki/Home/.

Korfmann K, Gaggiotti O, Fumagalli M. Deep learning in population genetics. *Genome Biol Evol*. 2023:15(2):evad008. https://doi.org/10.1093/gbe/evad008.

Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*. 2014:15(1):1–13. https://doi.org/10.1186/s12859-014-0356-4.

Krizhevsky A. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012:25:1097–1105.

Kruzel-Davila E, Wasser W, Skorecki K. APOL1 nephropathy: a population genetics and evolutionary medicine detective story. In: *Semin nephrol*. Vol. 37. Philadelphia (PA): Elsevier; 2017. p. 490–507.

Kuhlwilm M. The evolution of FOXP2 in the light of admixture. *Curr Opin Behav Sci*. 2018:21:120–126. https://doi.org/10.1016/j.cobeha.2018.04.006.

Labrie V, Buske O, Oh E, Jeremian R, Ptak C, Gasiunas G, Maleckas A, Petereit R, Žvirbliene A, Adamonis K, *et al*. Lactase nonpersistence is directed by DNA-variation-dependent epigenetic aging. *Nat Struct Mol Biol*. 2016:23(6):566–573. https://doi.org/10.1038/nsmb.3227.

Lauterbur M, Munch K, Enard D. Versatile detection of diverse selective sweeps with flex-sweep. *Mol Biol Evol*. 2023:40(6):msad139. https://doi.org/10.1093/molbev/msad139.

LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998:86(11):2278–2324. https://doi.org/10.1109/5.726791.

Lee K, Coop G. Distinguishing among modes of convergent adaptation using population genomic data. *Genetics*. 2017:207(4):1591–1619. https://doi.org/10.1534/genetics.117.300417.

Li C, Zhan Y, Xu C, Xu W, Wang S, Lv J, Zhou Y, Yue P, Chen B, Yang X. EDAG regulates the proliferation and differentiation of hematopoietic cells and resists cell apoptosis through the activation of nuclear factor-κb. *Cell Death Differ*. 2004:11(12):1299–1308. https://doi.org/10.1038/sj.cdd.4401490.

Lin Z, Li S, Feng C, Yang S, Wang H, Ma D, Zhang J, Gou M, Bu D, Zhang T, *et al.* Stabilizing mutations of KLHL24 ubiquitin ligase cause loss of keratin 14 and human skin fragility. *Nat Gen*. 2016:48(12):1508–1516. https://doi.org/10.1038/ng.3701.

Lloyd-Jones L, Zeng J, Sidorenko J, Yengo L, Moser G, Kemper K, Wang H, Zheng Z, Magi R, Esko T, *et al.* Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat Com*. 2019:10(1):5086. https://doi.org/10.1038/s41467-019-12653-0.

López-Bendito G, Flames N, Ma L, Fouquet C, Di Meglio T, Chedotal A, Tessier-Lavigne M, Marín O. Robo1 and Robo2 cooperate to control the guidance of major axonal tracts in the mammalian forebrain. *J Neurosci*. 2007:27(13):3395–3407. https://doi.org/10.1523/JNEUROSCI.4605-06.2007.

Lynch M. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci*. 2010:107(3):961–968. https://doi.org/10.1073/pnas.0912629107.

Masuda R, Kijima H, Imamura N, Aruga N, Nakazato K, Oiwa K, Nakano T, Watanabe H, Ikoma Y, Tanaka M, *et al.* Laminin-5γ2 chain expression is associated with tumor cell invasiveness and prognosis of lung squamous cell carcinoma. *Biomed Res*. 2012:33(5): 309–317. https://doi.org/10.2220/biomedres.33.309.

Mathieson S, Mathieson I. FADS1 and the timing of human adaptation to agriculture. *Mol Biol Evol*. 2018:35(12):2957–2970. https://doi.org/10.1093/molbev/msy180.

McArthur E, Rinker D, Capra J. Quantifying the contribution of neanderthal introgression to the heritability of complex traits. *Nat Commun*. 2021:12(1):4481. https://doi.org/10.1038/s41467-021-24582-y.

McManus K, Taravella A, Henn B, Bustamante C, Sikora M, Cornejo O. Population genetic analysis of the DARC locus (Duffy) reveals adaptation from standing variation associated with malaria resistance in humans. *PLoS Genet*. 2017:13(3):e1006560. https://doi.org/10.1371/journal.pgen.1006560.

McVicker G, Gordon D, Davis C, Green P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet*. 2009: 5(5):e1000471. https://doi.org/10.1371/journal.pgen.1000471.

Mignone F, Gissi C, Liuni S, Pesole G. Untranslated regions of mRNAs. *Genome Biol*. 2002:3(3):1–10. https://doi.org/10.1186/gb-2002-3-3-reviews0004.

Ming W, Cao C, Zhang G, Zhang H, Zhang F, Jiang Z, Yuan J. Application of convolutional neural network in defect detection of 3C products. *IEEE Access*. 2021:9:135657–135674. https://doi.org/10.1109/ACCESS.2021.3116131.

Miricescu D, Totan A, Stanescu-Spinu I, Badoiu SC, Stefani C, Greabu M. PI3K/AKT/mTOR signaling pathway in breast cancer: from molecular landscape to clinical aspects. *Int J Mol Sci*. 2020:22(1):173. https://doi.org/10.3390/ijms22010173.

Mughal M, Koch H, Huang J, Chiaromonte F, DeGiorgio M. Learning the properties of adaptive regions with functional data analysis. *PLoS Genet*. 2020:16(8):e1008896. https://doi.org/10.1371/journal.pgen.1008896.

Mughal MR, DeGiorgio M. Localizing and classifying adaptive targets with trend filtered regression. *Mol Biol Evol*. 2019:36(2):252–270. https://doi.org/10.1093/molbev/msy205.

Nair V, Hinton G. 2010. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10); Haifa, Israel. p. 807–814.

Nguyen C, Okamura T, Morita K, Yamaguchi S, Harada H, Miki Y, Izumo T, Kayamori K, Yamaguchi A, Sakamoto K. LAMC2 is a predictive marker for the malignant progression of leukoplakia. *J Oral Pathol Med*. 2017:46(3):223–231. https://doi.org/10.1111/jop.2017.46.issue-3.

Nicolaisen L, Desai M. Distortions in genealogies due to purifying selection and recombination. *Genetics*. 2013:195(1):221–230. https://doi.org/10.1534/genetics.113.152983.

Nielsen R, Paul J, Albrechtsen A, Song Y. Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet*. 2011:12(6):443–451. https://doi.org/10.1038/nrg2986.

Nielsen R, Williamson S, Kim Y, Hubisz M, Clark A, Bustamante C. Genomic scans for selective sweeps using SNP data. *Genome Res*. 2005:15(11):1566–1575. https://doi.org/10.1101/gr.4252305.

Oleksyk T, Smith M, O'Brien S. Genome-wide scans for footprints of natural selection. *Philos Trans R Soc Lond B Biol Sci*. 2010:365(1537):185–205. https://doi.org/10.1098/rstb.2009.0219.

Orang A, Dredge B, Liu C, Bracken J, Chen C, Sourdin L, Whitfield H, Lumb R, Boyle S, Davis M, *et al.* Basonuclin-2 regulates extracellular matrix production and degradation. *Life Sci Alliance*. 2023:6(10):e202301984. https://doi.org/10.26508/lsa.202301984.

Park Y, Ryu J. Models of synaptotagmin-1 to trigger CA^{2+}-dependent vesicle fusion. *FEBS Lett*. 2018:592(21):3480–3492. https://doi.org/10.1002/feb2.2018.592.issue-21.

Pasaniuc B, Zaitlen N, Shi H, Bhatia G, Gusev A, Pickrell J, Hirschhorn J, Strachan D, Patterson N, Price A. Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics*. 2014:30(20):2906–2914. https://doi.org/10.1093/bioinformatics/btu416.

Payseur B, Nachman M. Microsatellite variation and recombination rate in the human genome. *Genetics*. 2000:156(3):1285–1298. https://doi.org/10.1093/genetics/156.3.1285.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, *et al.* Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011:12: 2825–2830. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf.

Pickrell J, Coop G, Novembre J, Kudaravalli S, Li J, Absher D, Srinivasan B, Barsh G, Myers R, Feldman M, *et al.* Signals of recent positive selection in a worldwide sample of human populations. *Genome Res*. 2009:19(5):826–837. https://doi.org/10.1101/gr.087577.108.

Pierron D, Razafindrazaka H, Pagani L, Ricaut F, Antao T, Capredon M, Sambo C, Radimilahy C, Rakotoarisoa J, Blench R, *et al.* Genome-wide evidence of Austronesian–Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. *Proc Natl Acad Sci*. 2014:111(3):936–941. https://doi.org/10.1073/pnas.1321860111.

Racimo F. Testing for ancient selection using cross-population allele frequency differentiation. *Genetics*. 2016:202(2):733–750. https://doi.org/10.1534/genetics.115.178095.

Racimo F, Marnetto D, Huerta-Sánchez E. Signatures of archaic adaptive introgression in present-day human populations. *Mol Biol Evol*. 2017:34:296–317. https://doi.org/10.1093/molbev/msw216.

Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. Evidence for archaic adaptive introgression in humans. *Nat Rev Genet*. 2015:16(6):359–371. https://doi.org/10.1038/nrg3936.

Rogers A, Huff C. Linkage disequilibrium between loci with unknown phase. *Genetics*. 2009:182(3):839–844. https://doi.org/10.1534/genetics.108.093153.

Rudman N, Kaur S, Simunović V, Kifer D, Šoić D, Keser T, Štambuk T, Klarić L, Pociot F, Morahan G, *et al.* Integrated glycomics and genetics analyses reveal a potential role for N-glycosylation of plasma proteins and IgGS, as well as the complement system, in the development of type 1 diabetes. *Diabetologia*. 2023:66(6):1071–1083. https://doi.org/10.1007/s00125-023-05881-z.

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, *et al.* Imagenet large scale visual recognition challenge. *Int J Comput Vis*. 2015:115(3):211–252. https://doi.org/10.1007/s11263-015-0816-y.

Sabeti P, Reich D, Higgins J, Levine HZP, Richter D, Schaffner S, Gabriel S, Platko J, Patterson N, McDonald G, *et al.* Detecting recent positive selection in the human genome from haplotype structure.

Nature. 2002:419(6909):832–837. https://doi.org/10.1038/nature01140.

Sakharkar M, Chow VT, Kangueane P. Distributions of exons and introns in the human genome. *In Silico Biol*. 2004:4(4):387–393. https://doi.org/10.3233/ISB-00142.

Saxton R, Sabatini D. mTOR signaling in growth, metabolism, and disease. *Cell*. 2017:168(6):960–976. https://doi.org/10.1016/j.cell.2017.02.004.

Scally A, Durbin R. Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet*. 2012:13(10):745. https://doi.org/10.1038/nrg3295.

Schrider D. Background selection does not mimic the patterns of genetic diversity produced by selective sweeps. *Genetics*. 2020:216(2):499–519. https://doi.org/10.1534/genetics.120.303469.

Schrider D, Kern A. S/HIC: robust identification of soft and hard sweeps using machine learning. *PLoS Genet*. 2016:12(3):e1005928. https://doi.org/10.1371/journal.pgen.1005928.

Schrider D, Kern A. Soft sweeps are the dominant mode of adaptation in the human genome. *Mol Biol Evol*. 2017:34(8):1863–1877. https://doi.org/10.1093/molbev/msx154.

Schrider D, Kern A. Supervised machine learning for population genetics: a new paradigm. *Trends Genet*. 2018:34(4):301–312. https://doi.org/10.1016/j.tig.2017.12.005.

Schrider D, Mendes F, Hahn M, Kern A. Soft shoulders ahead: spurious signatures of soft and partial selective sweeps result from linked hard sweeps. *Genetics*. 2015:200(1):267–284. https://doi.org/10.1534/genetics.115.174912.

Seger J, Smith W, Perry J, Hunn J, Kaliszewska Z, Sala L, Pozzi L, Rowntree V, Adler F. Gene genealogies strongly distorted by weakly interfering mutations in constant environments. *Genetics*. 2010:184(2):529–545. https://doi.org/10.1534/genetics.109.103556.

Ségurel L, Bon C. On the evolution of lactase persistence in humans. *Ann Rev Genomics Hum Genet*. 2017:18(1):297–319. https://doi.org/10.1146/genom.2017.18.issue-1.

Selvaraju R, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. 2017. Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos (CA): IEEE Computer Society. p. 618–626.

Setter D, Mousset S, Cheng X, Nielsen R, DeGiorgio M, Hermisson J. VolcanoFinder: genomic scans for adaptive introgression. *PLoS Genet*. 2020:16(6):e1008867. https://doi.org/10.1371/journal.pgen.1008867.

Shafique A, Cao G, Khan Z, Asad M, Aslam M. Deep learning-based change detection in remote sensing images: a review. *Remote Sens*. 2022:14(4):871. https://doi.org/10.3390/rs14040871.

Shankarkumar U. The human leukocyte antigen (HLA) system. *Int J Hum Genet*. 2004:4(2):91–103. https://doi.org/10.1080/09723757.2004.11885875.

Sharif Razavian A, Azizpour H, Sullivan J, Carlsson S. 2014. CNN features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Los Alamitos (CA): IEEE Computer Society. p. 806–813.

Shibata H, Aramaki T, Sakai M, Ninomiya H, Tashiro N, Iwata N, Ozaki N, Fukumaki Y. Association study of polymorphisms in the GluR7, KA1 and KA2 kainate receptor genes (GRIK3, GRIK4, GRIK5) with schizophrenia. *Psychiatr Res*. 2006:141(1):39–51. https://doi.org/10.1016/j.psychres.2005.07.015.

Simonyan K, Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015). Computational and Biological Learning Society.

Slatkin M. Gene flow and the geographic structure of natural populations. *Science*. 1987:236(4803):787–792. https://doi.org/10.1126/science.3576198.

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014:15:1929–1958. https://jmlr.org/papers/v15/srivastava14a.html.

Stephens M, Balding D. Bayesian statistical methods for genetic association studies. *Nat Rev Genet*. 2009:10(10):681–690. https://doi.org/10.1038/nrg2615.

Stern A, Wilton P, Nielsen R. An approximate full-likelihood method for inferring selection and allele frequency trajectories from DNA sequence data. *PLoS Genet*. 2019:15(9):e1008384. https://doi.org/10.1371/journal.pgen.1008384.

Stipp C. Laminin-binding integrins and their tetraspanin partners as potential antimetastatic targets. *Expert Rev Mol Med*. 2010:12:e3. https://doi.org/10.1017/S1462399409001355.

Sturm RA, Larsson M. Genetics of human iris colour and patterns. *Pigment Cell Melanoma Res*. 2009:22(5):544–562. https://doi.org/10.1111/pcr.2009.22.issue-5.

Suganyadevi S, Seethalakshmi V, Balasamy K. A review on deep learning in medical image analysis. *Int J Multimed Inf Retr*. 2022:11(1):19–38. https://doi.org/10.1007/s13735-021-00218-1.

Sulem P, Gudbjartsson D, Stacey S, Helgason A, Rafnar T, Magnusson K, Manolescu A, Karason A, Palsson A, Thorleifsson G, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet*. 2007:39(12):1443–1452. https://doi.org/10.1038/ng.2007.13.

Szegedy C, Ioffe S, Vanhoucke V, Alemi A. 2017. Inception-v4, inception-ResNet and the impact of residual connections on learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. Washington (DC): Association for the Advancement of Artificial Intelligence (AAAI). Vol. 31.

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos (CA): IEEE Computer Society. p. 1–9.

Szpak M, Xue Y, Ayub Q, Tyler-Smith C. How well do we understand the basis of classic selective sweeps in humans? *FEBS Lett*. 2019:593(13):1431–1448. https://doi.org/10.1002/feb2.2019.593.issue-13.

Tajbakhsh N, Shin J, Gurudu S, Hurst R, Kendall C, Gotway M, Liang J. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging*. 2016:35(5):1299–1312. https://doi.org/10.1109/TMI.2016.2535302.

Tajima F. Evolutionary relationship of DNA sequences in finite populations. *Genetics*. 1983:105(2):437–460. https://doi.org/10.1093/genetics/105.2.437.

Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989:123(3):585–595. https://doi.org/10.1093/genetics/123.3.585.

Takei N, Nawa H. mTOR signaling and its roles in normal and abnormal brain development. *Front Mol Neurosci*. 2014:7:28. https://doi.org/10.3389/fnmol.2014.00028.

Talkowski M, Ernst C, Heilbut A, Hanscom CCC, Lindgren A, Kirby A, Liu S, Muddukrishna B, Ohsumi T, Shen Y, et al. Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. *Am J Hum Genet*. 2011:88(4):469–481. https://doi.org/10.1016/j.ajhg.2011.03.013.

Tan M, Le Q. 2019. Efficientnet: rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. Brookline (MA): Proceedings of Machine Learning Research (PMLR). p. 6105–6114.

Tennessen J, Bigham A, O'connor T, Fu W, Kenny E, Gravel S, McGee S, Do R, Liu X, Jun G, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 2012:337(6090):64–69. https://doi.org/10.1126/science.1219240.

The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015:526(7571):68–74. https://doi.org/10.1038/nature15393.

Tishkoff S, Reed F, Ranciaro A, Voight B, Babbitt C, Silverman J, Powell K, Mortensen H, Hirbo J, Osman M, et al. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet*. 2007:39(1):31–40. https://doi.org/10.1038/ng1946.

Torada L, Lorenzon L, Beddis A, Isildak U, Pattini L, Mathieson S, Fumagalli M. Imagene: a convolutional neural network to quantify natural selection from genomic data. *BMC Bioinformatics*. 2019:20(S9):337. https://doi.org/10.1186/s12859-019-2927-x.

Vaughn A, Nielsen R. Fast and accurate estimation of selection coefficients and allele histories from ancient and modern DNA. *Mol Biol Evol*. 2024:41(8):msae156. https://doi.org/10.1093/molbev/msae156.

Visser M, Palstra R, Kayser M. Human skin color is influenced by an intergenic DNA polymorphism regulating transcription of the nearby BNC2 pigmentation gene. *Hum Mol Genet*. 2014:23(21):5750–5762. https://doi.org/10.1093/hmg/ddu289.

Vitti J, Grossman S, Sabeti P. Detecting natural selection in genomic data. *Annu Rev Genet*. 2013:47(1):97–120. https://doi.org/10.1146/genet.2013.47.issue-1.

Voight B, Kudaravalli S, Wen X, Pritchard J. A map of recent positive selection in the human genome. *PLoS Biol*. 2006:4(3):e72. https://doi.org/10.1371/journal.pbio.0040072.

Vy H, Kim Y. A composite-likelihood method for detecting incomplete selective sweep from population genomic data. *Genetics*. 2015:200(2):633–649. https://doi.org/10.1534/genetics.115.175380.

Wang J, He X, Faming S, Lu G, Cong H, Jiang Q. A real-time bridge crack detection method based on an improved inception-ResNet-v2 structure. *IEEE Access*. 2021:9:93209–93223. https://doi.org/10.1109/ACCESS.2021.3093210.

Watterson G. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*. 1975:7(2):256–276. https://doi.org/10.1016/0040-5809(75)90020-9.

Wilde S, Timpson A, Kirsanow K, Kaiser E, Kayser M, Unterländer M, Hollfelder N, Potekhina I, Schier W, Thomas M, *et al*. Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc Natl Acad Sci USA*. 2014:111(13):4832–4837. https://doi.org/10.1073/pnas.1316513111.

Wu Y, Zhang X, Liu Y, Lu F, Chen X. Decreased expression of BNC1 and BNC2 is associated with genetic or epigenetic regulation in hepatocellular carcinoma. *Int J Mol Sci*. 2016:17(2):153. https://doi.org/10.3390/ijms17020153.

Xiao W, Wang R, Handy D, Loscalzo J. NAD(H) and NADP(H) redox couples and cellular energy metabolism. *Antioxid Redox Signal*. 2018:28(3):251–272. https://doi.org/10.1089/ars.2017.7216.

Xu J, Pang Z, Shin O, Südhof T. Synaptotagmin-1 functions as a CA^{2+} sensor for spontaneous release. *Nat Neurosci*. 2009:12(6):759–766. https://doi.org/10.1038/nn.2320.

Xue A, Schrider D, Kern A. Discovery of ongoing selective sweeps within anopheles mosquito populations using deep learning. *Mol Biol Evol*. 2021:38(3):1168–1183. https://doi.org/10.1093/molbev/msaa259.

Yang Y, Quan L, Ling Y. RBMS3 inhibits the proliferation and metastasis of breast cancer cells. *Oncol Res*. 2018:26(1):9. https://doi.org/10.3727/096504017X14871200709504.

Yilmaz M, Christofori G. EMT, the cytoskeleton, and cancer cell invasion. *Cancer Metastasis Rev*. 2009:28(1–2):15–33. https://doi.org/10.1007/s10555-008-9169-0.

Yin F, Sancheti H, Cadenas E. Silencing of nicotinamide nucleotide transhydrogenase impairs cellular redox homeostasis and energy metabolism in PC12 cells. *Biochim Biophys Acta Bioenerg*. 2012:1817(3):401–409. https://doi.org/10.1016/j.bbabio.2011.12.004.

Yin Q, Srivastava K, Gebremedhin A, Makuria A, Flegel W. Long-range haplotype analysis of the malaria parasite receptor gene ACKR1 in an East-African population. *Hum Genome Var*. 2018:5(1):26. https://doi.org/10.1038/s41439-018-0024-8.

Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? *Adv Neural Inf Process Syst*. 2014:27:3320–3328. https://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks.

Yu F, Koltun V. 2015. Multi-scale context aggregation by dilated convolutions. In: Presented at the International Conference on Learning Representations (ICLR) 2016. https://arxiv.org/abs/1511.07122.

Yu F, Seff A, Zhang Y, Song S, Funkhouser T, Xiao J. 2015. Lsun: construction of a large-scale image dataset using deep learning with humans in the loop, arXiv, arXiv:1506.03365, preprint: not peer reviewed.

Zhai Y, Shah M. 2006. Visual attention detection in video sequences using spatiotemporal cues. In: Proceedings of the 14th ACM International Conference on Multimedia. New York (NY): Association for Computing Machinery (ACM). p. 815–824.

Zhao H, Pavlidis P, Alachiotis N. 2023. SweepNet: a lightweight CNN architecture for the classification of adaptive genomic regions. In: Proceedings of the Platform for Advanced Scientific Computing Conference. New York (NY): Association for Computing Machinery (ACM). p. 1–10.

Zhao S, Zhou B, Luo F, Mao X, Lu Y. The structure and function of NKAIN2-a candidate tumor suppressor. *Int J Clin Exp Med*. 2015:8:17072.

Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. 2016. Learning deep features for discriminative localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos (CA): IEEE Computer Society. p. 2921–2929.

Zoncu R, Efeyan A, Sabatini D. mTOR: from growth signal integration to cancer, diabetes and ageing. *Nat Rev Mol Cel Biol*. 2011:12(1):21–35. https://doi.org/10.1038/nrm3025.

Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc B*. 2005:67(2):301–332. https://doi.org/10.1111/j.1467-9868.2005.00503.x.

Zuo L, Lu L, Tan Y, Pan X, Cai Y, Wang X, Hong J, Zhong C, Wang F, Zhang X, *et al*. Genome-wide association discoveries of alcohol dependence. *Am J Addict*. 2014:23(6):526–539. https://doi.org/10.1111/ajad.v23.6.

Zuo L, Tan Y, Zhang X, Wang X, Krystal J, Tabakoff B, Zhong C, Luo X. A new genomewide association meta-analysis of alcohol dependence. *Alcohol Clin Exp Res*. 2015:39(8):1388–1395. https://doi.org/10.1111/acer.12786.

Zuo L, Wang K, Zhang X, Li C, Zhang F, Wang X, Chen W, Gao G, Zhang H, Krystal J, *et al*. Rare SERINC2 variants are specific for alcohol dependence in subjects of European descent. *Pharmacogenet Genom*. 2013:23(8):395. https://doi.org/10.1097/FPC.0b013e328362f9f2.