LYCEUM: Learning to call copy number variants on low coverage ancient genomes

Mehmet Alper Yılmaz[1†], Ahmet Arda Ceylan[1†], Gün Kaynar[2†], and A. Ercüment Çiçek[1‡]

1. Department of Computer Engineering, Bilkent University, Ankara, Turkey
2. Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA
† Equal contribution, ‡ correspondence: cicek@cs.bilkent.edu.tr

Abstract. Copy number variants (CNVs) are pivotal in driving phenotypic variation that facilitates species adaptation. They are significant contributors to various disorders, making ancient genomes crucial for uncovering the genetic origins of disease susceptibility across populations. However, detecting CNVs in ancient DNA samples poses substantial challenges due to several factors. Ancient DNA (aDNA) is often highly degraded, and this degradation is further complicated by contamination from microbial DNA and DNA from closely related species, introducing additional noise into sequencing data. Finally, the typically low coverage of aDNA renders accurate CNV detection particularly difficult. Conventional CNV calling algorithms, optimized for high coverage and long reads, often underperform in such conditions. To address these limitations, we introduce LYCEUM, a deep learning-based CNV caller specifically designed for low-coverage aDNA. LYCEUM performs transfer learning from a model designed to detect CNVs in another noisy data domain, whole exome sequencing then it performs fine-tuning with a few aDNA samples for which semi-ground truth CNV calls are available. Our findings demonstrate that LYCEUM accurately identifies CNVs even in highly downsampled genomes, maintaining robust performance across a range of coverage levels. Thus, LYCEUM offers researchers a reliable solution for CNV detection in challenging ancient genomic datasets. LYCEUM is available at https://github.com/ciceklab/LYCEUM

Keywords: CNV detection · Ancient DNA · Deep Learning

1 Introduction

The analysis of ancient DNA (aDNA) offers valuable insights into species evolution, human migration, and historical population dynamics. With advancements in next-generation sequencing technologies, aDNA studies have evolved to reconstruct entire genomes, trace disease spread, and explore human ancestry, thus providing a more comprehensive understanding of genetic variation over time [1]. Despite these advancements, much of the genetic variation analysis in ancient DNA samples has predominantly focused on single nucleotide polymorphisms (SNPs), while the analysis of Copy Number Variations (CNVs) remains equally critical for understanding phenotypic diversity.

Gene duplication and deletion variants often exert a more substantial impact on phenotypes compared to point mutations and are key drivers of phenotypic variation that enable species to adapt to their environments [2–6]. For example, in human evolution, it has been shown that the human-specific copy number expansion of TCAF genes occurred approximately 1.7 million years ago, likely facilitating adaptation to cold climates or dietary shifts [7]. Furthermore, the TBC1D3 gene family, which plays a role in modulating epidermal growth factor receptor (EGFR) signaling, underwent significant expansion after the divergence of humans from chimpanzees [8]. This expansion is associated with the development of the human prefrontal cortex, underscoring the potential contribution of CNVs to human brain evolution. Collectively, these findings underscore the pivotal role of CNVs in human evolutionary processes. Additionally, CNV analysis in ancient DNA samples offers the potential to illuminate the historical prevalence of risk factors for complex diseases associated with CNVs. Recent research has highlighted the contribution of CNVs to susceptibility to complex traits and conditions, such as autism spectrum disorder (ASD) [9, 10], schizophrenia (SZ) [11], and bipolar disorder [12]. A recent CNV genome-wide association study (GWAS) identified 73 signals associated with 40 different diseases [13], all of which increase the risk for their respective conditions. Despite the well-documented role of CNVs in disease susceptibility, limited research has focused on the origins of

these variants. Thus, examining CNVs across diverse ancestral groups may offer insights into the origins of common disease susceptibilities in modern populations.

However, detecting structural variants (SVs), including CNVs, in ancient DNA samples poses considerable challenges. aDNA typically exhibits extensive postmortem damage, including deamination and fragmentation [14]. Moreover, pervasive microbial contamination often results in low endogenous DNA content, further complicated by contamination from closely related species, which makes it difficult to distinguish endogenous DNA from contaminant sequences [15, 16]. As a result, the endogenous DNA in ancient DNA samples often has low sequencing coverage, which complicates SV discovery. Many state-of-the-art whole genome sequencing (WGS)-based CNV callers rely on higher coverage [17–19] and long-read sequencing data [20, 21], making these approaches suboptimal for low-coverage ancient DNA samples.

To address these challenges, we introduce LYCEUM, a deep learning-based approach for detecting copy number variations (CNVs) in low-coverage ancient DNA samples. Due to the scarcity of high-coverage aDNA samples, it remains challenging to train a machine learning model for CNV detection, particularly a deep learning model. We propose that transfer learning from a model trained on a different noisy domain can facilitate CNV detection in aDNA data. To this end, we leverage transfer learning from a deep learning model designed to accurately detect CNVs in whole exome sequencing data, which is characterized by its notoriously noisy read depth signal. Subsequently, we fine-tune the model using high-coverage aDNA data with comparatively reliable CNV call labels to achieve high-performance CNV calling on low-coverage and quality aDNA data. We evaluate our model using simulated samples, moderate-coverage real ancient DNA samples, and its down-sampled versions. Remarkably, LYCEUM maintains the performance achieved on full-coverage datasets when applied to their down-sampled versions and outperforms existing archespores. Additionally, we demonstrate that the diversity patterns of the model's deletion calls in ancient DNA samples align with expected trends. Given its robustness in CNV calling and its adaptability for other ancient DNA samples through fine-tuning, we think LYCEUM will revolutionize CNV detection and genomic analyses of low-coverage ancient datasets in archaeogenetics and paleogenomics fields. LYCEUM is available at https://github.com/ciceklab/LYCEUM.

2 Results

2.1 Model Overview

Our model, LYCEUM builds upon the whole exome sequencing (WES)-based CNV caller, ECOLE [22]. This model was trained on close to 550 matched WES and whole genome sequencing (WGS) samples in the 1k Genomes project. Relying on a transformer-based architecture, it inputs noisy WES read depth signal and learns to predict per-exon CNV calls which are made using the reliable WGS read depth signal. LYCEUM is designed to transfer and configure this model to analyze low-coverage and noisy read depth signal at base-pair resolution within exon regions. LYCEUM first maps the aDNA WGS read-depth signal of each exon into a higher dimensional space and using a cascade of transformer encoders learns the likelihood of a CNV in the genomic context using a chromosome-specific classification token and positional encoding. The token is used for predicting the CNV event for each exon via fully connected layers. The system model is shown in Figure 1A. The model has been fine-tuned using moderate-coverage (>9x) aDNA samples, five down-sampled versions of these samples, and low-coverage ($\leq$ 1x) simulated aDNA samples. CNVnator[18] calls on original (high) coverage samples are used as semi-ground truth for real aDNA data.

2.2 Evaluation criteria

LYCEUM classifies each exon as a deletion, duplication, or no-call. So we report exon-level performance (precision, recall, and F1 score). For comparison, we map CNV calls from compared methods to the target exon regions and label each corresponding exon accordingly. In addition to this, we also report gene-level performances. To determine the overall CNV label for a gene, we aggregate the predicted CNV labels of all corresponding exons and assign a final gene-level CNV label using a majority voting approach. This means that the most frequent CNV label (deletion or duplication) among the exons is selected as the final gene-level CNV label (See Figure 1B). To ensure a fair comparison of performance across different models, we standardize this process by applying the same majority voting methodology to all models being compared.

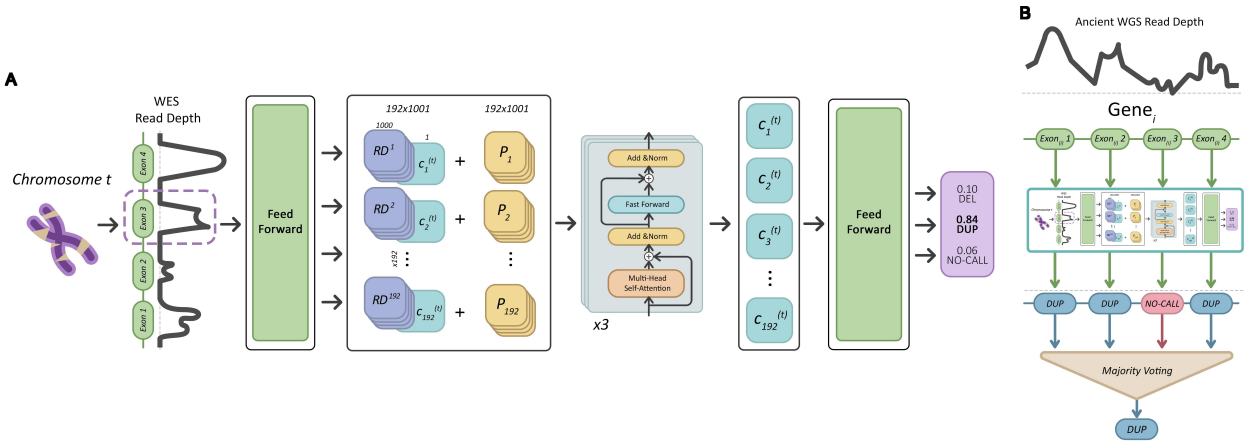

Fig. 1: The system model of LYCEUM. **(A)** Pretrained exome sequencing-based CNV caller model is shown. The read depth (RD) for each exon (1kbp) is mapped to a higher dimensional space (192x1000) at base-pair resolution and concatenated with a chromosome-specific classification token (c) (192x1) and summed up with positional encoding (P, 192x1001). This tensor is passed through 3 transformer encoder blocks and the sliced classification token is input to a fully connected neural network which outputs the call probabilities. **(B)** LYCEUM uses the WGS-based aDNA read depth signal that overlaps with the exons to make per-exon calls. We use the CNVnator calls as labels, obtained on simulated samples and moderate to high coverage aDNA samples (>9x). To obtain gene-level calls majority voting on all exons of a gene is performed.

First, the predicted CNV calls from each model were mapped to the target exon regions. After aligning the exon-level predictions, we determine the gene-level CNV label for each model using the same majority voting strategy.

2.3 The performance of LYCEUM on low-coverage ancient DNA samples

To evaluate the performance of LYCEUM on low-coverage ancient DNA samples, we use 20 aDNA samples for which we have moderate coverage (>9x) obtained from eight published studies [23–30]. Since these samples have relatively higher coverage compared to a typical aDNA which has <1x coverage, we treat them as regular WGS samples and employ the WGS CNV caller CNVnator [18] to obtain semi-ground truth CNV calls. To assess the robustness of the models at varying depths of sequencing coverage, we down-sample samples to specific coverage levels—5x, 1x, 0.5x, 0.1x, and 0.05x—using samtools-view [31]. This step allows us to simulate the challenges posed by low-coverage ancient genomes and observe how well the models maintain performance under such conditions. We use 13 samples in the training set and test the model on the remaining 7 samples. For benchmarking, we compare LYCEUM against the only available aDNA CNV caller CONGA [32] and the state-of-the-art WGS CNV callers CONTROL-FREEC [19] and CNVnator. Note that there is no ground truth CNV call set for aDNA samples. Our goal with LYCEUM is to be able to call CNVs on low-coverage aDNA data with similar performance to CNV calling on moderate-to-high coverage WGS data.

Figure 2 shows the F1 score comparison of the methods for detecting deletions and duplications in genes. We also present the exon-level and gene-level precision and recall results in Supplementary Figures 1 and 2, respectively, along with the corresponding confusion matrices in Supplementary Tables 3-10. Note that CNVnator's F1 score on the original coverage is 1.0 since since these results are used as the semi-ground truth calls to measure performance. Yet, as the coverage goes down its performance consistently deteriorates. CONGA, on the other hand, achieves very high precision but very low recall using the suggested hyperparameters (see Supplementary Tables 3-10) and its performance also deteriorates with lower coverage. On the other hand, LYCEUM demonstrates consistent performance across all coverage levels for both deletions and duplications, maintaining its performance even at very low coverages (0.1x and 0.05x), where it outperforms other tools. We also observe that it maintains higher recall as opposed to CONGA, close to the level of a

regular WGS-based CNV caller. It achieves a 17.5% and 38.9% F1 score improvement over the next best method at the lowest coverage (0.05x) for deletions and duplications, respectively. This highlights LYCEUM's robustness in low-coverage conditions, a critical advantage in ancient genome analysis where such scenarios are common.

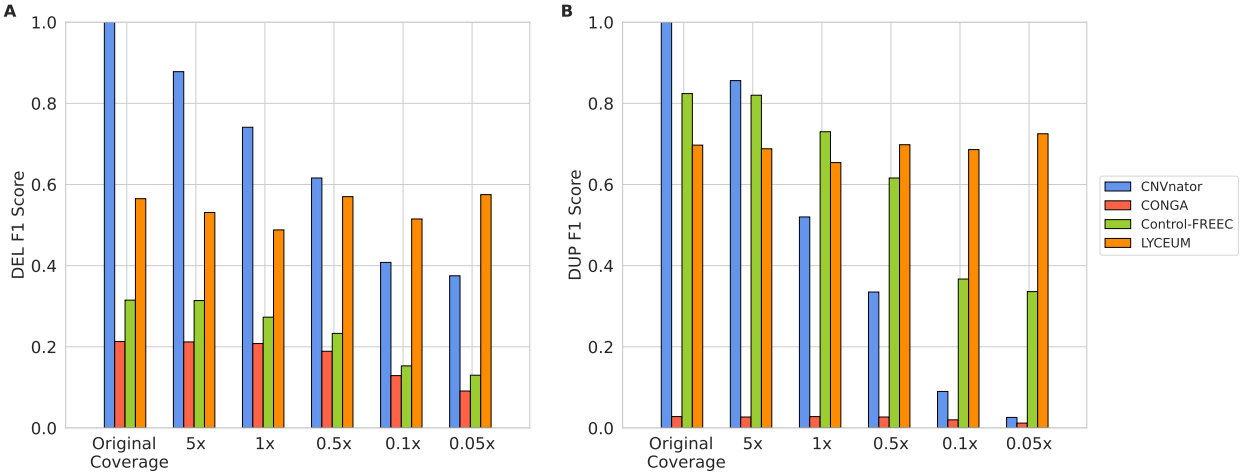

Fig. 2: A comparative analysis of F1 scores for CNV caller algorithms applied to seven ancient genome samples with varying coverage levels across gene regions. CNVnator calls on the original coverage to serve as a semi-ground truth reference. Each sample is down-sampled to five different coverage levels, and each algorithm is subsequently tested on these samples. **(A)** F1 score results for deletion detection. **(B)** F1 score results for duplication detection.

2.4 Comparative performance analysis on simulated ancient genomic data

We also compare the methods using simulated aDNA data. We use the VarSim tool [33] to insert structural variations, specifically duplications and deletions observed in the samples from the 1000 Genomes Project [14]. We incorporate these to the reference human genome to create a realistic range of genetic changes. We generate 15 different genomes which contain duplications and deletions that range from 500 bps to 100 kbps. Each simulated genome has approximately 2,000 duplications and 2,000 deletions, capturing a wide spectrum of genetic alterations. To mimic the characteristics of aDNA, the genomes generated by VarSim are further processed with the Gargammel tool [34] which simulates post-mortem DNA damage [14]. This involves fragmenting the DNA into shorter sequences and adding degradation patterns typical of ancient genomes (See Section 4.1 for more details on simulation). To assess the robustness of the models in challenging conditions, we generate 60 simulated ancient genome samples at 4 different coverage levels: —0.05x, 0.1x, 0.5x, and 1x— out of 15 VarSim-generated genomes. It simulates the noisy and incomplete nature of ancient genomes. We then apply LYCEUM and other compared methods to the resulting samples and obtain gene-level CNV labels

Figure 3 shows the results. Detailed exon- and gene-level performance scores are provided in Supplementary Tables 11, and 12, along with corresponding confusion matrices in Supplementary Tables 13-20. We observe that LYCEUM exhibits superior performance compared to other methods at the lowest coverage value (0.05x), achieving F1 score improvements of 65.8% and 71.2% over the second-best method for deletions and duplications, respectively. Additionally, LYCEUM 's performance remains highly consistent with decreasing coverage, emphasizing its robustness. This suggests that the LYCEUM's transformer architecture is particularly effective for scenarios involving highly degraded and incomplete datasets, where conventional tools often struggle.

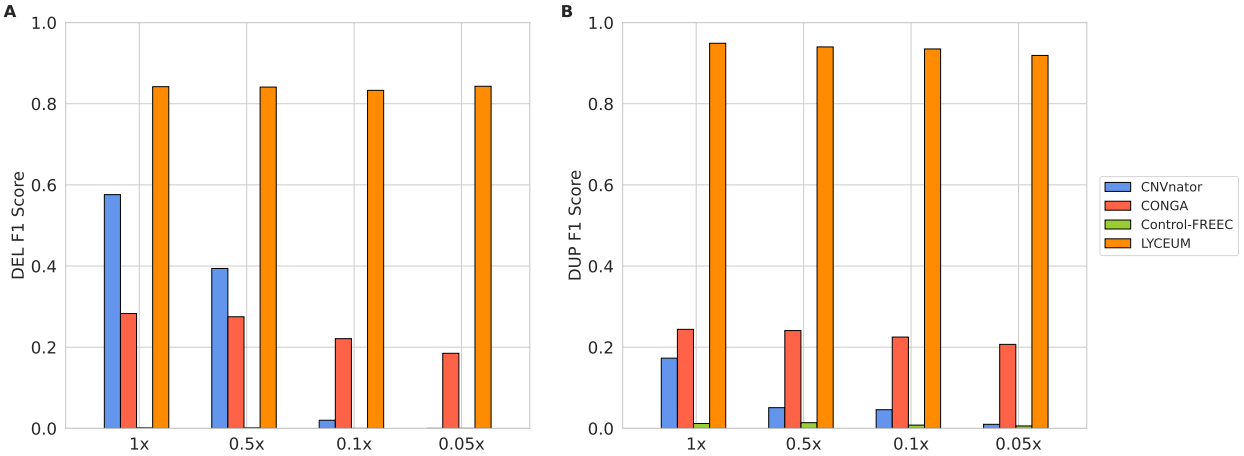

Fig. 3: Comparison of F1 scores among CNV caller algorithms evaluated on simulated ancient genomes with differing coverage levels in gene regions. **(A)** F1 score performance for detecting deletions. **(B)** F1 score performance for detecting duplications.

2.5 Segmental deletions detected by LYCEUM recover the demographic history

To assess the reliability of LYCEUM, we analyze the distribution of high-confidence (probability > 0.9) and variable (variance > 0.2) exon-level deletion calls across 50 ancient DNA samples. The geographic locations of the findings are shown in Figure 4A where darker blue indicates western Europe, and darker red indicates eastern Asia. We apply Principal Component Analysis (PCA) to visualize the genetic background of these samples in a 2D space, shown in Figure 4B. Notably, the PCA plot reveals two primary clusters, corresponding to samples from Europe and Asia, aligning with the geographic distribution. We observe that samples mostly cluster among west and east origins and the samples from central Asia are distributed across two clusters. Additionally, we compare the clustering performances of LYCEUM with CONGA. Figure 4C presents the PCA plot of the deletion events detected by CONGA with variance > 0.2, across the same 50 ancient DNA samples. The subsequent calculation of the silhouette index reveals that both LYCEUM and CONGA yield the same index: 0.11. Furthermore, hierarchical clustering of exon deletion calls from both CONGA and LYCEUM demonstrates again clustering accuracy of 72% (See Figure 5). The clustering patterns further validate the capability of LYCEUM to capture meaningful demographic patterns using genetic structure in low-coverage ancient DNA samples.

2.6 Segmental deletions are negatively selected

Deletion events, particularly as copy number variations, often adversely impact fitness. These deletion variants are subject to negative selection and are gradually eliminated from the gene pool over time [35, 36]. Moreover, longer deletions are observed less frequently due to stronger negative selection, likely because they have more significant functional consequences [37]. In our analysis shown in Figure 6 we use 80,896 autosomal deletion events detected called by LYCEUM on 50 ancient DNA samples to prove that LYCEUM's calls are biologically consistent and align with known evolutionary dynamics.

The Site Frequency Spectrum (SFS) in Figure 6A illustrates the relationship between mean allele frequency and the density of deletions. The figure shows that deletions with lower allele frequencies have higher densities, indicating that less frequent deletions are much more dense in the autosomal deletion call set. This pattern is consistent with the expected influence of negative selection, where rare deletions are maintained at low frequencies due to selective pressure against them [35].

The results in Figure 6B indicate that longer deletions are substantially less frequent. This observation is consistent with the expectation that longer deletions face stronger negative selection pressures [37]. Specifically, we found a strong negative correlation (Spearman correlation = -0.92) between deletion size and density. It demonstrates that natural selection acts more intensively against larger deletions, which leads to their lower representation in the population.

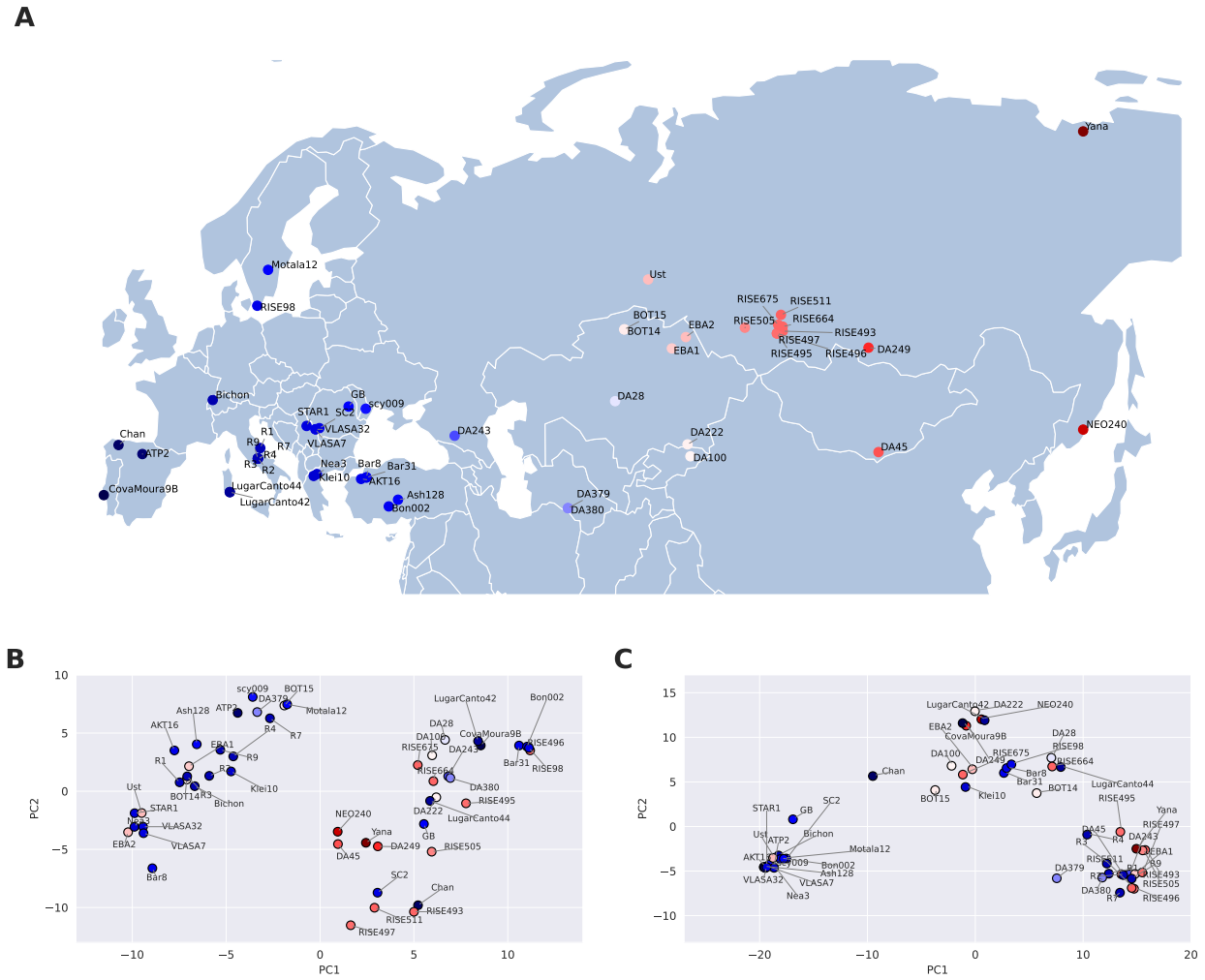

Fig. 4: Distribution of segmental deletions (**A**) Geographic locations of 50 ancient DNA samples. The color spectrum represents the relative positions of samples from east (red) to west (blue). (**B**) and (**C**) show the visualization of the first two principal components representing genomic distances based on segmental deletion events identified by LYCEUM and CONGA, respectively. Results highly correlate with demographic structure

3 Discussion

The results of this study reveal the efficacy of LYCEUM, our deep learning-based CNV caller, in addressing the unique challenges associated with ancient DNA analysis, particularly in detecting copy number variations from low-coverage and noisy genome data. As demonstrated, LYCEUM maintains high performance across varying depths of sequencing coverage, substantially outperforming other CNV callers such as CONTROL-FREEC, CNVnator, and also the state-of-the-art ancient genome CNV caller CONGA, especially in minimal coverage. We also show that the detected CNV composition can accurately recover the geographical origins of the samples and reconfirm that the deletion events are under negative selection.

Despite these promising results, some limitations remain. The reliance on CNVnator calls as semi-ground truth for real ancient DNA samples, while necessary due to the lack of true ground truth, might be introducing potential biases into performance evaluation. But we aim to validate this by comparing LYCEUM and CNVnator calls on lower coverage configurations. Future work should aim to validate LYCEUM's predictions using independently verified CNV datasets from ancient genomes as they become available.

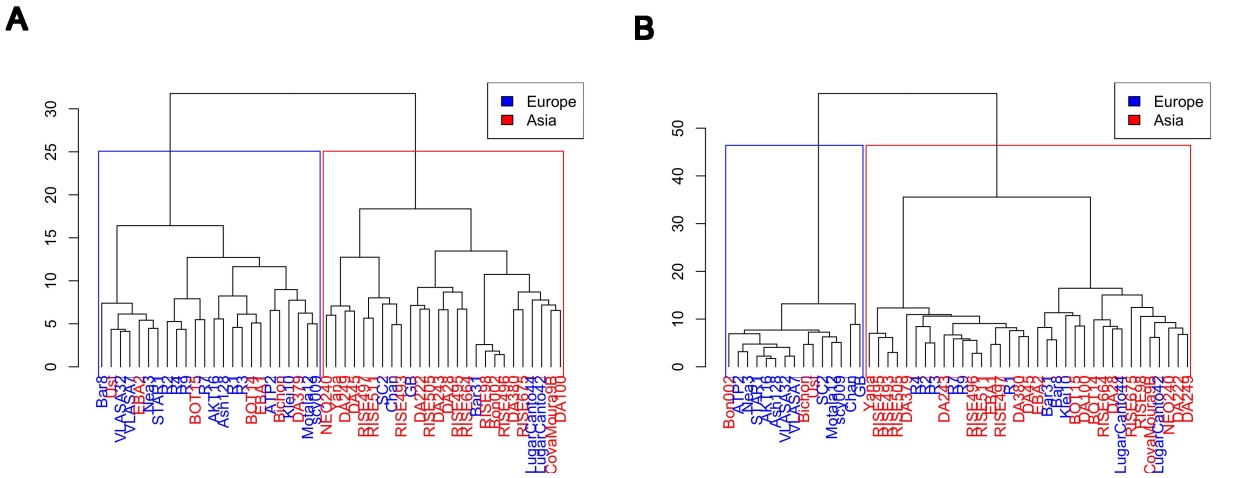

Fig. 5: Hierarchical clustering dendrograms of LYCEUM (**A**) and CONGA (**B**) on 50 aDNA samples. Both dendrograms demonstrate a similar clustering trend according to geographic origin, highlighting genetic distinctions between ancient European and Asian populations.

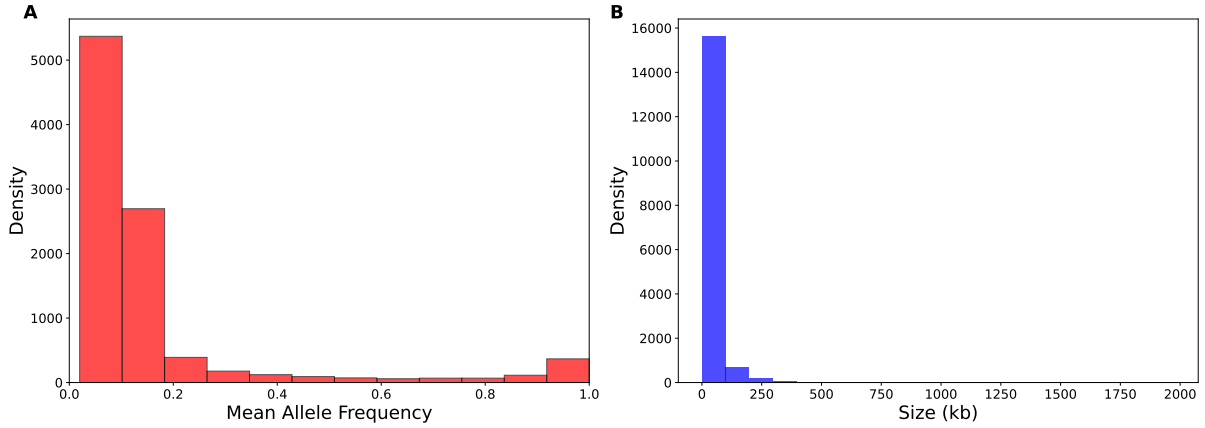

Fig. 6: Impact of negative selection autosomal deletion variants. **(A)** Analysis of the relationship between the mean allele frequency of these deletions and their density, illustrating how allele frequency influences the prevalence of deletion variants across the genome. **(B)** Examination of the relationship between deletion size and the density of these deletions, providing insights into how the size of deletion variants correlates with their distribution frequency within autosomal regions.

In conclusion, LYCEUM represents a significant advancement in the field of ancient DNA analysis, offering a powerful tool for CNV detection in low-coverage and highly degraded datasets. Its ability to maintain accuracy across different coverage levels, combined with its scalability through fine-tuning strategy, makes LYCEUM a valuable resource for researchers studying evolutionary processes, human migration, and the genetic underpinnings of disease susceptibility in ancient populations.

4 Methods

4.1 Data Sets

Ancient Genome Dataset We use 63 aDNA WGS samples from West and East Eurasia, as well as North America [23–30, 38–48]. We present the metadata for these datasets in Supplementary Tables 21-23. Sample

coverage levels vary widely, ranging from 0.04 to 26.3 (median = 4.2). To get the reliable ground truth, we separate 20 moderate coverage samples (>9x) and run CNVnator on these samples. We obtain 31,946 DUP and 55,337 DEL calls in total at the exon level and 3,324 DUP and 6,823 DEL calls at the gene level. True negative samples are the remaining exons or genes for which CNVnator has not made any calls for exon-specific and gene-specific settings, respectively. These calls collectively form the semi-ground truth CNV calls. Then, we down-sample each sample into 5x, 1x, 0.5x, 0.1x and 0.05x coverages. Among the 20 samples, we randomly choose 7 samples (together with corresponding down-sampled samples) in the dataset and assign them to the test set. We use "Ust", "VLASA7", "Bichon", "VLASA32", "AKT16", "Nea3", and "STAR1" as test samples and the rest of the dataset (13 samples) as training samples. The test samples are used to evaluate all models. (see Supplementary Tables 21-22). Specifically, the training samples in the dataset are labeled with 18,967 DUP and 47,795 DEL calls in total for the exon-specific setting and, with 2,019 DUP and 5,730 DEL calls for the gene-specific setting. On the other hand, test samples in the dataset are labeled with 12,979 DUP and 7,542 DEL calls for the exon-specific setting and, with 1,305 DUP and 1,093 DEL calls for the gene-specific setting.

For the analyses described in Section 2.5 and 2.6, We use the remaining 43 samples and the 7 test samples mentioned above (Supplementary 23). The coverage values of these 50 samples are between 0.04 and 26.3 (median = 3.5) Note that we do not include down-sampled version of these 7 samples in these analyses.

Ancient genome simulation. We use the VarSim [33] tool to insert DEL and DUP variations into the reference genome hg38. This method is a variant simulation framework used for both germline and somatic mutations. We randomly select 15 WGS samples from the 1000 Genomes Project. We then input these 15 labels to the VarSim tool with the following parameters: sv_num_del = 2000, sv_num_dup = 2000, sv_percent_novel = 0.01, sv_min_length_lim = 500, sv_max_length_lim = 100000 and total_coverage = 10. These parameters allow us to simulate novel DEL and DUP variations while preserving a genome's variational distribution. The simulated 15 genomes contain DEL and DUP insertions in the size range of 500 bps and 100 kbps. Next, we use the Gargammel [14] tool to simulate ancient genomes. This tool is specifically developed to generate artificial ancient genome fragments. We input the 15 simulated genomes with DEL and DUP insertions to Gargammel with the following parameters: nick frequency = 0.024, length of overhanging ends = 0.36, probability of deamination of Cs in double-stranded parts = 0.0097, and probability of deamination of Cs in single-stranded parts = 0.55. These parameters allow us to simulate post-mortem DNA damage, including fragmenting DNA into shorter sequences and adding degradation patterns typical for ancient DNA samples. We run Gargammel with 1x, 0.5x, 0.1x, and 0.05x coverages for each Varsim output, in order to further mimic low-coverage ancient DNA. We randomly pick 5 Varsim generated genomes and use their corresponding 20 Gargammel-processed versions for training and we use the remaining 40 samples as test samples. All of the 60 samples in the dataset are labeled with 43,335 DUP and 21,469 DEL calls for the exon-specific setting and, with 7,313 DUP and 4,459 DEL calls for the gene-specific setting. True negative samples are the remaining exons or genes for which CNVnator has not made any calls for exon-specific and gene-specific settings respectively.

4.2 Problem Formulation

The problem is to decide if an exon i in an ancient DNA sample, has a CNV (i.e., DEL or DUP) or not. We denote exon i using X^i. Each exon is associated with (i) a chromosome (X^i_{ch}), (ii) coordinates in that chromosome: X^i_{start} and X^i_{end}, and finally, (iii) a normalized read depth signal vector: X^i_{RD} which can be at most 1 kbp-long and padded from right, if shorter. Longer exons are discarded which are relatively very rare. Y^i is the semi-ground truth label for exon i, obtained using CNVnator tool on the original full coverage read depth signal for real and simulated aDNA samples.

Let $ECOLE(X^i, \theta)$ be the ECOLE model which is a multi-class classifier that performs exon-wise CNV prediction (i.e DEL, DUP, NO-CALL) using the WES-based read depth signal, parameterized by θ. LYCEUM fine-tunes this model using the aDNA WGS read depth signal to obtain θ'. $\hat{Y^i} = LYCEUM(X^i, \theta')$ is the prediction of the LYCEUM model for the exon i and the model's parameters θ' is optimized to minimize the difference between $\hat{Y^i}$ and the semi-ground truth labels Y^i obtained using CNVnator. To obtain gene-level calls, we perform majority voting among all exons of a gene.

4.3 Model Description

We show the architecture of LYCEUM in Figure 1. The model first inputs the read dept signal $X^i_{RD} \in \mathbb{R}^{1000\times1}$ and maps it to a higher dimension $\mathbf{X}^i_{RD} \in \mathbb{R}^{192\times1000}$ using a fully connected network. A classification token $c^k \in \mathbb{R}^{192}$, which is specific to each chromosome k, is appended to $\mathbf{X}^i_{RD}$ to obtain the final input tensor $\hat{\mathbf{X}}^i_{RD} \in \mathbb{R}^{192\times1001}$ which will be input to the cascade of transformer encoder blocks. This classification token contains learnable parameters by the model and is used to classify the label for this exon in the context of this chromosome, later in the model. Finally, $\hat{\mathbf{X}}^i_{RD}$ is summed up with a same-sized positional encoding tensor to inform the model about the exact position of each read depth value embedding within chromosome k. This tensor is passed through a cascade of three transformer encoding blocks to obtain the transformed output of the same size $\mathbf{O}^i_{RD} \in \mathbb{R}^{192\times1001}$. The last column of the tensor that contains the chromosome-specific token is sliced $\mathbf{O}^i_{RD}[:, 1001] \in \mathbb{R}^{192\times1}$ and passed through a fully connected network followed by softmax to predict DEL, DUP and NO-CALL probabilities for exon i. $\hat{Y^i}$ is assigned as the class which has the maximum probability among these three options.

Approximately 5% of ancient exon targets may lack read depth information which prohibits LYCEUM to make any calls for those exons. We address this issue by applying a majority voting for that exon system based on $\hat{Y^i}$ of the three nearest neighbor exons.

4.4 Experimental Setup

Compared Methods We compare LYCEUM against following WGS-based CNV callers: CNVnator[18], CONGA[19] and Control-FREEC[32]. We use CNVnator's CNV calls on full-coverage data as semi-ground truth labels and test CNVnator and others on down-sampled data. LYCEUM, CNVnator, and CONGA report categorical CNV predictions. Control-FREEC reports integer CNVs. To include this method in our comparison, we discretize its predictions. We use DEL if the tool's output is less than 2, DUP if it is larger than 2 and NO-CALL if it is 2.

Parameter Settings For CNVnator, we use the recommended bin size parameter based on coverage in their paper. Specifically, we use bin size 1,000 for full coverage, 3,000 for 5x, 15000 for 1x, 30,000 for 0.5x, 150,000 for 0.1x, and 300,000 for 0.05x data. For CONGA and Control-FREEC, all parameters are set to their default values.

Training LYCEUM We use the pre-trained base ECOLE model from [22]. This model was trained using WES read depth signal of close to 550 randomly selected 1000 Genomes samples as input, and CNV predictions were obtained on matched WGS samples as labels. We fine-tune it using (i) WGS-based 13 aDNA training samples, (ii)five down-sampled versions of each of the samples in (i) (5x, 1x, 0.5x, 0.1x, 0.05x), and 20 simulated ancient DNA samples at 4 coverage levels (1x, 0.5x, 0.1x, 0.05x). The read depth signals overlapping with exon regions are used as input, with CNV calls from the original full-coverage aDNA data generated by CNVnator serving as labels for real samples, while the output VCF file from VarSim provides the ground truth for simulated samples. We name this model LYCEUM. During fine-tuning, we use Adam optimizer [49] and cosine annealing learning rate scheduler with an initial learning rate of $5x10^{-5}$. The model converged in 2 epochs.

We train all models on a SuperMicro SuperServer 4029GP-TRT with 2 Intel Xeon Gold 6140 Processors (2.3GHz, 24.75M cache) and 256GB RAM. We used a single NVIDIA GeForce RTX 2080 Ti GPU (24GB, 384Bit). Fine-tuning took approximately 72 hours. Note that users do not need to train a model from scratch and can use the released LYCEUM model for inference which is rapid. The average time to call all CNVs per ancient genome is $\sim$ 5 mins.

Data Availability

LYCEUM is implemented and released at https://github.com/ciceklab/LYCEUM under CC-BY-NC-ND 4.0 International license. The samples, ground truth data, and CNV predictions for real and simulated datasets to reproduce the analyses can be accessed through https://zenodo.org/records/13997011.

Acknowledgments

We would like thank Alperen Utku Yalçın, Erfan Farhang Kia, Dilek Koptekin, Can Alkan and Mehmet Somel for their valuable contributions to this paper.

References

1. Ludovic Orlando, Robin Allaby, Pontus Skoglund, Clio Der Sarkissian, Philipp W Stockhammer, María C Ávila-Arcos, Qiaomei Fu, Johannes Krause, Eske Willerslev, Anne C Stone, and Christina Warinner. Ancient DNA analysis. *Nat. Rev. Methods Primers*, 1(1), February 2021.
2. Jackson Peter, Matteo De Chiara, Anne Friedrich, Jia-Xing Yue, David Pflieger, Anders Bergström, Anastasie Sigwalt, Benjamin Barre, Kelle Freel, Agnès Llored, et al. Genome evolution across 1,011 saccharomyces cerevisiae isolates. *Nature*, 556(7701):339–344, 2018.
3. Claire Mérot, Rebekah A Oomen, Anna Tigano, and Maren Wellenreuther. A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology & Evolution*, 35(7):561–572, 2020.
4. Maren Wellenreuther, Claire Mérot, Emma Berdan, and Louis Bernatchez. Going beyond snps: The role of structural genomic variants in adaptive evolution and species diversification. *Molecular ecology*, 28(6), 2019.
5. Luzia Stalder, Ursula Oggenfuss, Norfarhan Mohd-Assaad, and Daniel Croll. The population genetics of adaptation through copy number variation in a fungal plant pathogen. *Molecular Ecology*, 32(10):2443–2460, 2023.
6. Wolfram Höps, Tobias Rausch, Michael Jendrusch, Jan O Korbel, and Fritz J Sedlazeck. Impact and characterization of serial structural variations across humans and great apes. *Nature Communications*, 15(1):8007, 2024.
7. Pinghsun Hsieh, Vy Dang, Mitchell R Vollger, Yafei Mao, Tzu-Hsueh Huang, Philip C Dishuck, Carl Baker, Stuart Cantsilieris, Alexandra P Lewis, Katherine M Munson, Melanie Sorensen, Annemarie E Welch, Jason G Underwood, and Evan E Eichler. Evidence for opposing selective forces operating on human-specific duplicated TCAF genes in neanderthals and humans. *Nature Communications*, 12(1):5118, August 2021.
8. Mitchell R. Vollger, Xavi Guitart, Philip C. Dishuck, Ludovica Mercuri, William T. Harvey, Ariel Gershman, Mark Diekhans, Arvis Sulovari, Katherine M. Munson, Alexandra P. Lewis, Kendra Hoekzema, David Porubsky, Ruiyang Li, Sergey Nurk, Sergey Koren, Karen H. Miga, Adam M. Phillippy, Winston Timp, Mario Ventura, and Evan E. Eichler. Segmental duplications and their variation in a complete human genome. *Science*, 376(6588):eabj6965, 2022.
9. Hreinn Stefansson, Dan Rujescu, Sven Cichon, Olli P. Pietiläinen, Andres Ingason, Stacy Steinberg, Ragnheidur Fossdal, Engilbert Sigurdsson, Thordur Sigmundsson, Jacobine E. Buizer-Voskamp, and et al. Large recurrent microdeletions associated with schizophrenia. *Nature*, 455(7210):232–236, 2008.
10. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, 455(7210):237–241, 2008.
11. Christelle M Durand, Catalina Betancur, Tobias M Boeckers, Juergen Bockmann, Pauline Chaste, Fabien Fauchereau, Gudrun Nygren, Maria Rastam, I Carina Gillberg, Henrik Anckarsäter, and et al. Mutations in the gene encoding the synaptic scaffolding protein shank3 are associated with autism spectrum disorders. *Nature Genetics*, 39(1):25–27, 2006.
12. Dheeraj Malhotra and Jonathan Sebat. Cnvs: harbingers of a rare variant revolution in psychiatric genetics. *Cell*, 148(6):1223–1241, 2012.
13. Chiara Auwerx, Maarja Jõeloo, Marie C Sadler, Nicolò Tesio, Sven Ojavee, Charlie J Clark, Reedik Mägi, Estonian Biobank Research Team Esko Tõnu Metspalu Andres Milani Lili Nelis Mari, Alexandre Reymond, and Zoltán Kutalik. Rare copy-number variants as modulators of common disease susceptibility. *Genome Medicine*, 16(1):5, 2024.
14. Adrian W Briggs, Udo Stenzel, Philip LF Johnson, Richard E Green, Janet Kelso, Kay Prüfer, Matthias Meyer, Johannes Krause, Michael T Ronan, Michael Lachmann, et al. Patterns of damage in genomic dna sequences from a neandertal. *Proceedings of the National Academy of Sciences*, 104(37):14616–14621, 2007.
15. Bárbara Sousa da Mota, Simone Rubinacci, Diana Ivette Cruz Dávalos, Carlos Eduardo G. Amorim, Martin Sikora, Niels N Johannsen, Marzena H Szmyt, Piotr Włodarczak, Anita Szczepanek, Marcin M Przybyła, et al. Imputation of ancient human genomes. *Nature Communications*, 14(1):3660, 2023.
16. Stéphane Peyrégne and Kay Prüfer. Present-day dna contamination in ancient dna datasets. *Bioessays*, 42(9):2000081, 2020.
17. Johannes Smolander, Sofia Khan, Kalaimathy Singaravelu, Leni Kauko, Riikka J Lund, Asta Laiho, and Laura L Elo. Evaluation of tools for identifying large copy number variations from ultra-low-coverage whole-genome sequencing data. *BMC Genomics*, 22(1):357, May 2021.
18. Alexej Abyzov, Alexander E Urban, Michael Snyder, and Mark Gerstein. Cnvnator: an approach to discover, genotype, and characterize typical and atypical cnvs from family and population genome sequencing. *Genome research*, 21(6):974–984, 2011.

19. Valentina Boeva, Tatiana Popova, Kevin Bleakley, Pierre Chiche, Julie Cappo, Gudrun Schleiermacher, Isabelle Janoueix-Lerosey, Olivier Delattre, and Emmanuel Barillot. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*, 28(3):423–425, 12 2011.
20. Mark JP Chaisson, John Huddleston, Megan Y Dennis, Peter H Sudmant, Maika Malig, Fereydoun Hormozdiari, Francesca Antonacci, Urvashi Surti, Richard Sandstrom, Matthew Boitano, et al. Resolving the complexity of the human genome using single-molecule sequencing. *Nature*, 517(7536):608–611, 2015.
21. Fritz J Sedlazeck, Philipp Rescheneder, Moritz Smolka, Han Fang, Maria Nattestad, Arndt Von Haeseler, and Michael C Schatz. Accurate detection of complex structural variations using single-molecule sequencing. *Nature methods*, 15(6):461–468, 2018.
22. Berk Mandiracioglu, Furkan Ozden, Gun Kaynar, Mehmet Alper Yilmaz, Can Alkan, and A Ercument Cicek. Ecole: Learning to call copy number variants on whole exome sequencing data. *Nature Communications*, 15(1):132, 2024.
23. Morten Rasmussen, Sarah L Anzick, Michael R Waters, Pontus Skoglund, Michael DeGiorgio, Thomas W Stafford Jr, Simon Rasmussen, Ida Moltke, Anders Albrechtsen, Shane M Doyle, et al. The genome of a late pleistocene human from a clovis burial site in western montana. *Nature*, 506(7487):225–229, 2014.
24. Nina Marchi, Laura Winkelbach, Ilektra Schulz, Maxime Brami, Zuzana Hofmanová, Jens Blöcher, Carlos S Reyna-Blanco, Yoan Diekmann, Alexandre Thiéry, Adamandia Kapopoulou, et al. The mixed genetic origin of the first farmers of europe. *bioRxiv*, pages 2020–11, 2020.
25. Peter de Barros Damgaard, Nina Marchi, Simon Rasmussen, Michaël Peyrot, Gabriel Renaud, Thorfinn Korneliussen, J Víctor Moreno-Mayar, Mikkel Winther Pedersen, Amy Goldberg, Emma Usmanova, et al. 137 ancient human genomes from across the eurasian steppes. *Nature*, 557(7705):369–374, 2018.
26. Gülşah Merve Kılınç, Natalija Kashuba, Dilek Koptekin, Nora Bergfeldt, Handan Melike Dönertaş, Ricardo Rodríguez-Varela, Dmitrij Shergin, Grigorij Ivanov, Dmitrii Kichigin, Kjunnej Pestereva, et al. Human population dynamics and yersinia pestis in ancient northeast asia. *Science Advances*, 7(2):eabc4587, 2021.
27. Eppie R Jones, Gloria Gonzalez-Fortes, Sarah Connell, Veronika Siska, Anders Eriksson, Rui Martiniano, Russell L McLaughlin, Marcos Gallego Llorente, Lara M Cassidy, Cristina Gamba, et al. Upper palaeolithic genomes reveal deep roots of modern eurasians. *Nature communications*, 6(1):8912, 2015.
28. Iosif Lazaridis, Nick Patterson, Alissa Mittnik, Gabriel Renaud, Swapan Mallick, Karola Kirsanow, Peter H Sudmant, Joshua G Schraiber, Sergi Castellano, Mark Lipson, et al. Ancient human genomes suggest three ancestral populations for present-day europeans. *Nature*, 513(7518):409–413, 2014.
29. Farnaz Broushaki, Mark G Thomas, Vivian Link, Saioa López, Lucy Van Dorp, Karola Kirsanow, Zuzana Hofmanová, Yoan Diekmann, Lara M Cassidy, David Díez-del Molino, et al. Early neolithic genomes from the eastern fertile crescent. *Science*, 353(6298):499–503, 2016.
30. Qiaomei Fu, Heng Li, Priya Moorjani, Flora Jay, Sergey M Slepchenko, Aleksei A Bondarev, Philip LF Johnson, Ayinuer Aximu-Petri, Kay Prüfer, Cesare De Filippo, et al. Genome sequence of a 45,000-year-old modern human from western siberia. *Nature*, 514(7523):445–449, 2014.
31. Petr Danecek, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, Thomas Keane, Shane A McCarthy, Robert M Davies, and Heng Li. Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), 02 2021. giab008.
32. Arda Söylev, Sevim Seda Çokoglu, Dilek Koptekin, Can Alkan, and Mehmet Somel. Conga: Copy number variation genotyping in ancient genomes and low-coverage sequencing data. *PLOS Computational Biology*, 18(12):e1010788, 2022.
33. John C Mu, Marghoob Mohiyuddin, Jian Li, Narges Bani Asadi, Mark B Gerstein, Alexej Abyzov, Wing H Wong, and Hugo YK Lam. Varsim: a high-fidelity simulation and validation framework for high-throughput genome sequencing with cancer applications. *Bioinformatics*, 31(9):1469–1471, 2015.
34. Gabriel Renaud, Kristian Hanghøj, Eske Willerslev, and Ludovic Orlando. gargammel: a sequence simulator for ancient DNA. *Bioinformatics*, 33(4):577–579, 11 2016.
35. Gregory M Cooper, Deborah A Nickerson, and Evan E Eichler. Mutational and selective effects on copy-number variants in the human genome. *Nature genetics*, 39(Suppl 7):S22–S29, 2007.
36. Donald F Conrad and Matthew E Hurles. The population genetics of structural variation. *Nature genetics*, 39(Suppl 7):S30–S36, 2007.
37. Feng Zhang, Wenli Gu, Matthew E Hurles, and James R Lupski. Copy number variation in human health, disease, and evolution. *Annual review of genomics and human genetics*, 10(1):451–481, 2009.
38. Gülşah Merve Kılınç, Ayça Omrak, Füsun Özer, Torsten Günther, Ali Metin Büyükkarakaya, Erhan Bıçakçı, Douglas Baird, Handan Melike Dönertaş, Ayshin Ghalichi, Reyhan Yaka, et al. The demographic development of the first farmers in anatolia. *Current Biology*, 26(19):2659–2666, 2016.
39. Margaret L Antonio, Ziyue Gao, Hannah M Moots, Michaela Lucci, Francesca Candilio, Susanna Sawyer, Victoria Oberreiter, Diego Calderon, Katharina Devitofranceschi, Rachael C Aikens, et al. Ancient rome: A genetic crossroads of europe and the mediterranean. *Science*, 366(6466):708–714, 2019.

40. Torsten Günther, Cristina Valdiosera, Helena Malmström, Irene Ureña, Ricardo Rodriguez-Varela, Óddny Osk Sverrisdóttir, Evangelia A Daskalaki, Pontus Skoglund, Thijessen Naidoo, Emma M Svensson, et al. Ancient genomes link early farmers from atapuerca in spain to modern-day basques. *Proceedings of the National Academy of Sciences*, 112(38):11917–11922, 2015.
41. Z Hofmanová, S Kreutzer, G Hellenthal, C Sell, Y Diekmann, D Díez-Del-Molino, and J Burger. Early farmers from across europe directly descended from neolithic aegeans. pnas, 2016.
42. Martin Sikora, Vladimir V Pitulko, Vitor C Sousa, Morten E Allentoft, Lasse Vinner, Simon Rasmussen, Ashot Margaryan, Peter de Barros Damgaard, Constanza de la Fuente, Gabriel Renaud, et al. The population history of northeastern siberia since the pleistocene. *Nature*, 570(7760):182–188, 2019.
43. Morten E Allentoft, Martin Sikora, Karl-Göran Sjögren, Simon Rasmussen, Morten Rasmussen, Jesper Stenderup, Peter B Damgaard, Hannes Schroeder, Torbjörn Ahlström, Lasse Vinner, et al. Population genomics of bronze age eurasia. *Nature*, 522(7555):167–172, 2015.
44. Maja Krzewińska, Gülşah Merve Kılınç, Anna Juras, Dilek Koptekin, Maciej Chyleński, Alexey G Nikitin, Nikolai Shcherbakov, Iia Shuteleva, Tatiana Leonova, Liudmila Kraeva, et al. Ancient genomes suggest the eastern pontic-caspian steppe as the source of western iron age nomads. *Science advances*, 4(10):eaat4457, 2018.
45. Peter de Barros Damgaard, Rui Martiniano, Jack Kamm, J Víctor Moreno-Mayar, Guus Kroonen, Michaël Peyrot, Gojko Barjamovic, Simon Rasmussen, Claus Zacho, Nurbol Baimukhanov, et al. The first horse herders and the impact of early bronze age steppe expansions into asia. *Science*, 360(6396):eaar7711, 2018.
46. Gloria González-Fortes, Eppie R Jones, Emma Lightfoot, Clive Bonsall, Catalin Lazar, Aurora Grandal-d'Anglade, María Dolores Garralda, Labib Drak, Veronika Siska, Angela Simalcsik, et al. Paleogenomic evidence for multi-generational mixing between neolithic farmers and mesolithic hunter-gatherers in the lower danube basin. *Current Biology*, 27(12):1801–1810, 2017.
47. Reyhan Yaka, Igor Mapelli, Damla Kaptan, Ayça Doğu, Maciej Chyleński, Ömür Dilek Erdal, Dilek Koptekin, Kıvılcım Başak Vural, Alex Bayliss, Camilla Mazzucato, et al. Variable kinship patterns in neolithic anatolia revealed by ancient genomes. *Current Biology*, 31(11):2455–2468, 2021.
48. Rui Martiniano, Lara M Cassidy, Ros Ó'Maoldúin, Russell McLaughlin, Nuno M Silva, Licinio Manco, Daniel Fidalgo, Tania Pereira, Maria J Coelho, Miguel Serra, et al. The population genomics of archaeological transition in west iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLoS genetics*, 13(7):e1006852, 2017.
49. Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.