

BARCODE: A Chromosome Karyotype Analysis Model Based on a Banding Code Framework

Kaijie Wu

kaijiewu@sjtu.edu.cn

Shanghai Jiao Tong University

Zehan Qiu

Shanghai Jiao Tong University

Wenjing Hu

International Peace Maternity And Child Health Hospital

Hao Cheng

Shanghai Jiao Tong University

Jiawei Cao

Shanghai Jiao Tong University

Chuchu Gong

Shanghai Jiao Tong University

Shuyuan Li

International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University

Haodong Pan

Shanghai Jiao Tong University

Chendi Li

International Peace Maternity And Child Health Hospital

Xinping Guan

Shanghai Jiao Tong University <https://orcid.org/0000-0003-1858-8538>

Hong Xu

The International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai

Article

Keywords:

Posted Date: August 11th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10344242/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

BARCODE: A Chromosome Karyotype Analysis Model Based on a Banding Code Framework

Zehan Qiu^{1,2}, Wenjing Hu³, Hao Cheng^{1,2}, Jiawei Cao^{1,2}, Chuchu Gong^{1,2}, Shuyuan Li³, Haodong Pan^{1,2}, Chendi Li⁴, Xinping Guan^{1,2}, Hong Xu^{4*}, and Kaijie Wu^{1,2*}

¹School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, Shanghai, China

²Key Laboratory for System Control and Information Processing, Ministry of Education of China, Shanghai 200240

³Department of Reproductive Genetics, International Peace Maternity And Child Health Hospital, Shanghai, China

⁴Department of Gynecology, International Peace Maternity And Child Health Hospital, Shanghai, China

*Corresponding author. Email: kaijiewu@sjtu.edu.cn , xuhong1558@sjtu.edu.cn

Abstract

Chromosome karyotyping is an important clinical method for identifying chromosomal numerical abnormalities and structural variants. However, its automation has long been limited by the scarcity and heterogeneity of structurally abnormal samples, which constrain supervised learning across diverse rearrangements, and by the lack of interpretable representations needed for fine-grained band-level localization. Here, we present BARCODE, a banding-code-based image-to-sequence model for chromosome karyotype analysis. This method converts continuous G-banding grayscale patterns into biologically meaningful Banding Code sequences aligned with International System for Human Cytogenomic Nomenclature (ISCN2024) logic, and integrates a morphology-aware full-width visual embedding module, a prior-guided banding augmentation strategy, and an image-text matching mechanism, in which Banding Code sequences are treated as structured text-like representations. Through coding-sequence generation, BARCODE learns the intrinsic grammar of chromosome banding arrangements from normal chromosome samples. Without using real abnormal chromosomes during training, BARCODE enables zero-shot structural abnormality detection, robust chromosome classification under distribution shift, and interpretable band-level localization through explicit code deviations. Meanwhile, the image-code matching score supports image-quality stratification and uncertainty estimation. By transforming chromosome images into clinically meaningful symbolic sequences, this Banding Code framework provides an interpretable and data-efficient solution for intelligent cytogenetic analysis.

1 Introduction

Chromosome karyotyping remains an indispensable clinical approach for identifying chromosomal numerical abnormalities and structural variants. Balanced chromosomal rearrangements usually do not alter genomic dosage, and carriers may remain phenotypically normal; however, these rearrangements can lead to the formation of unbalanced gametes and are therefore an important contributor to infertility and recurrent miscarriage [1, 2]. Although molecular techniques such as chromosomal microarray analysis, copy-number variation sequencing, and whole-genome sequencing have advanced rapidly, they cannot fully replace banding-based karyotyping. By providing a genome-wide view of chromosome structure, conventional karyotyping remains particularly valuable for detecting balanced rearrangements, large-scale structural abnormalities, and low-level mosaicism that copy-number-based methods may fail to identify [3, 4, 5, 6, 7, 8, 9, 10, 11].

Despite its clinical importance, the interpretation of structural abnormalities still relies heavily on experienced cytogeneticists. According to the International System for Human Cytogenomic Nomenclature 2024 (ISCN2024), experts must manually inspect chromosome banding patterns, compare homologous chromosomes, and localize potential abnormal breakpoints [12, 13]. This process is time-consuming, difficult to scale, and susceptible to inter-observer variability, particularly when chromosome bands are blurred, staining is uneven, or structural rearrangements are complex. Therefore, achieving efficient and reliable structural abnormality analysis while preserving diagnostic sensitivity and interpretability remains a major challenge in clinical cytogenetics.

Deep learning offers a promising route toward automated chromosome analysis, yet its application to structural abnormality detection remains constrained by three closely related challenges [14, 15, 16, 17, 18]. First, real

structural abnormalities are rare and highly heterogeneous, resulting in a pronounced long-tail distribution and making it difficult to construct sufficiently large and representative training datasets. This scarcity is partly associated with biological selection against severe chromosomal aberrations, many of which lead to early embryonic lethality and are consequently underrepresented in clinically observable samples [19, 20, 21, 22]. Second, conventional supervised models are generally optimized for predefined abnormality categories represented in the training data. Their detection capability is therefore often limited to previously observed and explicitly annotated abnormal patterns, leading to poor generalization to unseen, rare, or combinatorial structural variants. Third, most existing methods rely primarily on statistical image features and do not explicitly represent the biological organization of chromosome bands [23, 14, 24]. Meanwhile, precise band-level analysis would typically require dense expert annotations that are costly, time-consuming, and difficult to standardize [25, 18, 26, 27, 28]. Together, these limitations restrict model generalization, interpretability, fine-grained localization, and compatibility with standardized cytogenetic reporting.

To address these challenges, we introduce Banding Code, a biologically grounded symbolic representation inspired by the G-band ideograms and topological logic of ISCN2024 [12, 13]. Chromosome G-banding patterns are not arbitrary grayscale textures but reflect macroscopic differences in DNA sequence composition, replication timing, gene density, chromatin organization, and staining response [29, 30, 31, 32]. Banding Code discretizes continuously varying chromosomal grayscale patterns into ordered symbolic sequences that preserve both band composition and longitudinal organization. Rather than treating a chromosome only as a pixel matrix, this representation encodes chromosome bands as biologically meaningful units and connects low-level visual information with high-level cytogenetic semantics. Consequently, the model can move beyond learning purely discriminative image features and instead capture banding regularities that can be directly interpreted using established cytogenetic knowledge.

Building on this representation, we develop BARCODE, an image-to-sequence model for chromosome karyotype analysis, drawing on advances in image-to-sequence generation and Transformer-based sequence modeling [33, 34, 35]. BARCODE learns the composition and longitudinal arrangement of chromosome bands and transforms visual banding features into structured sequences with explicit cytogenetic meaning. By training on normal chromosome samples together with biologically plausible banding perturbations generated from cytogenetic priors, the model avoids dependence on large collections of real structurally abnormal chromosomes and does not require dense pixel-level abnormality annotations. BARCODE therefore supports robust chromosome classification, zero-shot structural abnormality detection, and interpretable band-level localization within a unified framework. Deviations in the generated Banding Code sequences provide explicit evidence of abnormal banding regions, while image-code consistency scores further support image-quality stratification and uncertainty estimation.

In summary, BARCODE reframes automated karyotyping from conventional image classification into a biologically informed image-to-sequence generation task. By embedding chromosome banding knowledge into the representation itself, the framework jointly addresses abnormal-sample scarcity, the limited generalization of supervised models, and insufficient interpretability. Without using real structurally abnormal chromosomes during training, BARCODE enables sensitive structural abnormality screening and fine-grained band-level analysis. Its structured outputs may improve the efficiency and consistency of preliminary clinical screening and review, while also providing a foundation for the future generation of ISCN-compatible karyotype reports.

2 Results

Overview of BARCODE

BARCODE aims to enable standardized and intelligent chromosome karyotype analysis through the Banding Code. At its core, this method introduces the Banding Code as an intermediate representation that transforms continuous grayscale image information into discrete numerical sequences conforming to ISCN standards, thereby achieving fine-grained and consistent representation of chromosome banding features. The biological basis of this representation is grounded in the established relationship between G-banding patterns, DNA sequence composition, replication timing, gene density, and chromatin organization [29, 30, 31, 32]. The concept of the Banding Code can be generalized as a universal methodology applicable across diverse karyotyping frameworks. Our model accepts straightened chromosomes as input and leverages an image-to-Banding Code pre-training paradigm to generate discrete Banding Code sequences alongside final karyotype diagnostic results. First, BARCODE integrates a prior-guided band enhancement strategy, which equips the model with zero-shot recognition capabilities for unseen abnormal types without requiring real abnormal training samples. Second, a morphology-aware full-width visual embedding module is employed to extract embeddings of complete band structural features along the chromosome’s central axis. These embeddings are then processed by a Transformer-based encoder-decoder architecture to generate the Banding Code [35]. During training, we design a multi-task learning mechanism that jointly optimizes language modeling (LM) loss and image-code matching (ITM) loss, inspired by image-text alignment and image-text matching paradigms in vision-language learning [36, 34, 37]. This mechanism auto-

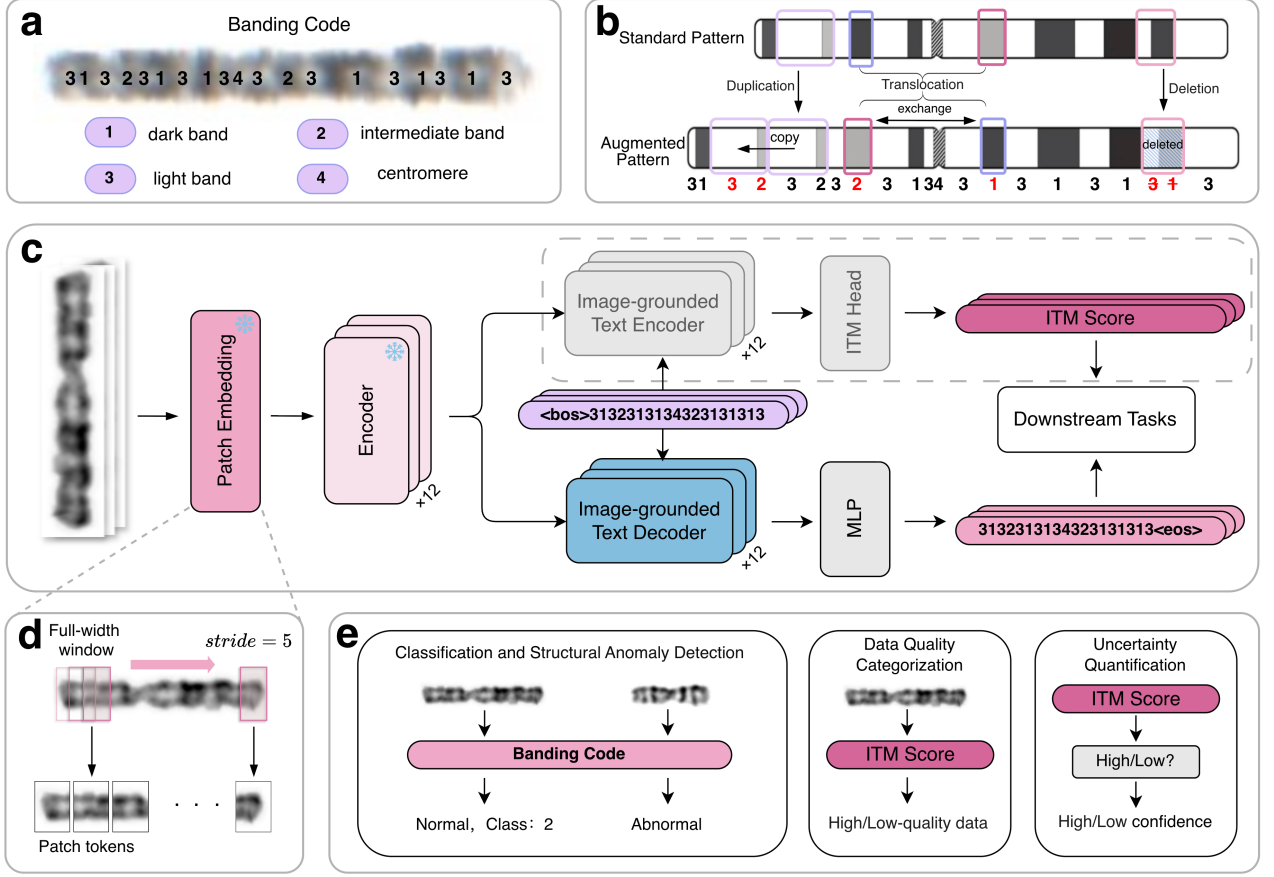


Figure 1: **Review of the BARCODE for chromosome karyotyping.** **a**, Illustration of the Banding Code representation. Each discrete token encodes biologically meaningful regions. **b**, Example of structural anomaly simulation through prior-guided banding augmentation. Standard chromosome patterns (top) are transformed via duplication, translocation, and deletion to generate augmented sequences (bottom), with corresponding Banding Code sequences indicated. **c**, Overall architecture of the BARCODE image-to-sequence model. During the two-stage training, modules within the dashed box are excluded, and those marked with snowflakes are frozen in the second stage. **d**, Detailed visualization of the morphology-aware full-width embedding. A full-width sliding window extracts longitudinal visual tokens from each chromosome, preserving complete cross-sectional banding information and sequence order. **e**, Downstream tasks supported by the BARCODE, including chromosome classification across different banding resolutions, structural abnormality detection, image quality categorization, and uncertainty quantification using ITM scores.

matically categorizes training data into high- and low-quality subsets, enabling a progressive training strategy. Furthermore, the ITM score output by the model serves as a prediction confidence metric, thereby strengthening decision reliability. BARCODE simultaneously supports multiple downstream tasks, including multi-resolution chromosome classification, structural abnormality detection, chromosome quality grading, and uncertainty quantification. An overview of BARCODE is summarized in Figure 1, with comprehensive details provided in the Methods section.

Exploring the consistency of BARCODE’s Banding Code generation

Existing chromosome analysis paradigms frequently overlook banding patterns with critical diagnostic value and instead rely on visual feature learning derived from underlying continuous grayscale distributions. As inherent biological characteristics, G-banding patterns encode diagnostic semantics aligned with ISCN standards and offer substantially greater clinical interpretability through discrete distribution patterns rather than continuous grayscale signals [3, 12, 13]. These banding patterns are not arbitrary visual textures, but reflect macro-scale differences in DNA composition, replication timing, gene density, and chromatin organization [29, 30, 31, 32]. To address this visual-semantic mismatch, we propose a novel karyotyping paradigm centered on the Banding Code. This paradigm explicitly maps the model’s analytical perspective from raw grayscale variations to biologically grounded banding structures. The concept originates from ISCN human chromosome G-banding ideograms, which systematically document canonical band distribution rules. Through discretization and sym-

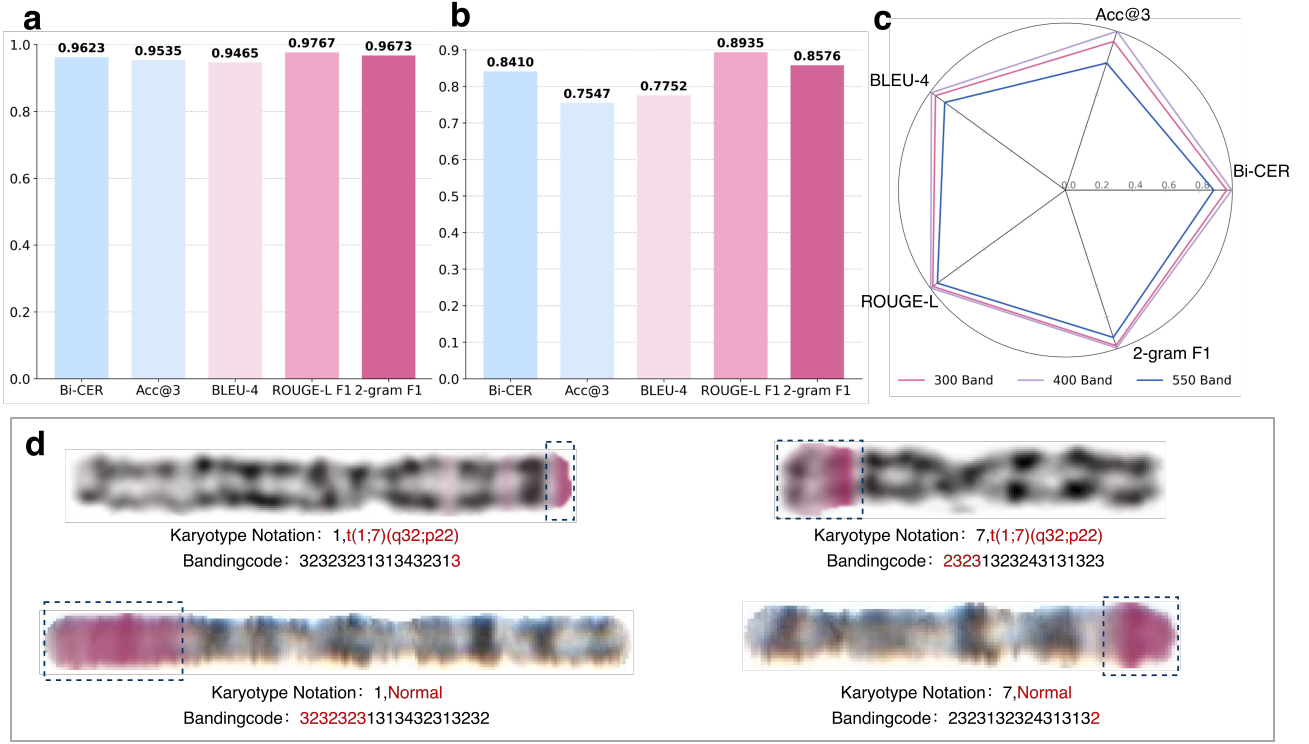


Figure 2: Evaluation of Banding Code generation consistency and interpretability. **a**, Performance of BARCODE on normal chromosome instances. **b**, Performance of BARCODE on structurally abnormal chromosome instances. **c**, Comparison of Banding Code generation performance across different banding resolutions. **d**, Representative examples of Banding Code generation and their spatial correspondence. Dashed boxes indicate chromosome regions associated with highlighted Banding Code tokens. Examples include structurally abnormal chromosomes with translocation annotations and normal chromosomes, demonstrating that deviations in the generated codes correspond to visually and biologically meaningful banding regions.

bolization of continuous grayscale patterns, we construct an ISCN-compliant sequential representation known as the Banding Code. This transformation fundamentally refines the model’s perceptual granularity to match clinical interpretation standards.

The fidelity of Banding Code generation serves as the cornerstone of reliable karyotyping and directly dictates the credibility of downstream diagnostic outcomes. We systematically evaluated the code generation capability of BARCODE across both normal chromosomes and structurally abnormal samples. This assessment verifies the model’s capacity to accurately capture banding patterns and produce high-fidelity sequences that adhere to established biological priors. We comprehensively evaluate model performance using multiple complementary metrics. ROUGE-L F1, 2-gram F1, and BLEU-4 quantify sequence generation quality, following widely used automatic sequence evaluation protocols [38, 39]. Bi-CER (Bidirectional Code Correctness Rate) measures exact code matching precision. Acc@3 calculates the proportion of samples with a bidirectional edit distance of no more than three, thereby characterizing model robustness under fault-tolerant conditions. Detailed training configurations and specific formulations for Bi-CER and Acc@3 are provided in the Methods section.

Figure 2a illustrates the Banding Code generation capability of BARCODE on normal chromosomes. Results demonstrate consistently high scores across standard sequence generation metrics. This performance confirms that the model accurately captures discrete band categories while precisely preserving the longitudinal topological arrangement between adjacent bands. Specialized metrics designed for the Banding Code, including Bi-CER and Acc@3, also achieve high performance levels. These results indicate that generated codes maintain robust structural integrity despite minor band boundary localization deviations, thereby preventing diagnostic misclassification. Furthermore, Figure 2b evaluates the Banding Code generation capability of BARCODE on structurally abnormal chromosomes. Despite exclusive exposure to normal chromosome samples during training, the model demonstrates exceptional zero-shot code generation performance. This outcome confirms that BARCODE transcends simple training data memorization and successfully learns the intrinsic biological grammar of chromosomal banding arrangements alongside the essential information of individual bands.

Figure 2c presents sequence generation experiments across datasets with varying banding resolutions. These results further validate the consistent code generation capability of BARCODE. This trend is consistent with the distribution of banding resolutions in routine clinical datasets, where 400-band preparations account for the majority of samples, whereas 550-band preparations are less represented. This trend closely correlates with

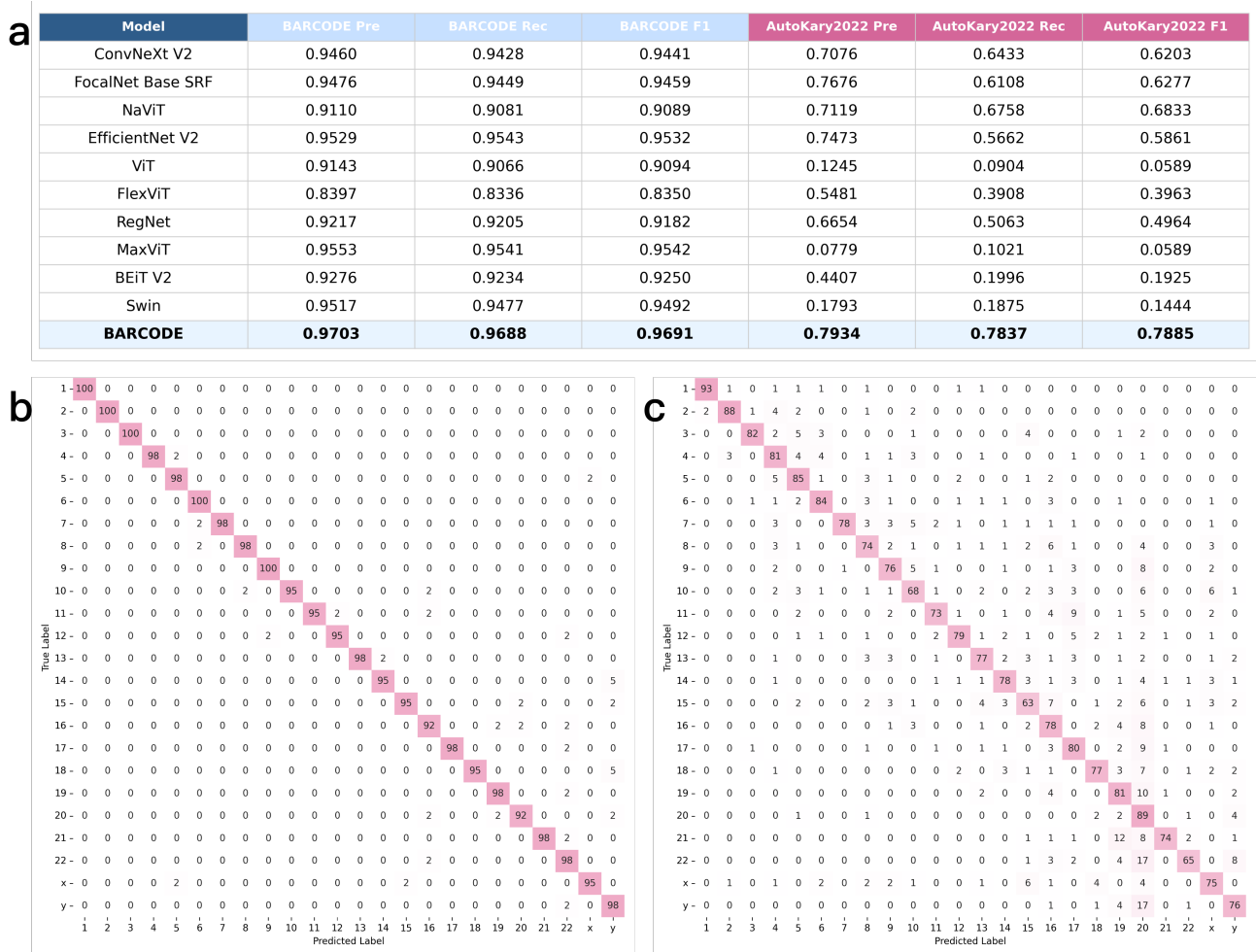


Figure 3: Performance comparison of BARCODE and baseline models for chromosome classification.
a, Comparison of performance metrics for BARCODE and baseline models on the internal validation dataset and the out-of-domain AutoKary2022 dataset. Metrics include Precision, Recall, and F1-score. BARCODE achieves the highest scores across all datasets, demonstrating both high classification performance and robustness under distribution shift. **b**, Confusion matrix for BARCODE on the internal validation dataset. **c**, Confusion matrix for BARCODE on the AutoKary2022 dataset.

banding-resolution distribution patterns in clinical datasets. Standard 400-band preparations dominate routine clinical practice while 550-band or higher samples remain relatively rare. Consequently, data-driven statistical priors for Banding Codes inevitably skew toward the dominant distribution. This distribution bias introduces minor generalization deviations when processing rare high-band-resolution samples. BARCODE effectively adapts to grayscale distributions across different banding resolutions. The generated Banding Code sequences maintain logical consistency across varying banding resolutions. This robust generalization capability directly ensures model applicability in real-world clinical workflows.

In contrast to conventional image-level deep learning models that rely on implicit visual representations, the BARCODE delivers intrinsic interpretability through explicit Banding Code sequences. We conducted a multi-dimensional interpretability analysis to validate the biological rationality of the model decision logic. Figure 2d visualizes the direct correspondence between generated codes and specific band positions. Cross-attention weights for specific abnormal codes concentrate precisely on actual structural aberration regions within chromosome images. Attention mapping for normal chromosome Banding Codes also demonstrates exact spatial consistency between highlighted model response regions and physical band coordinates. These findings confirm that BARCODE captures diagnostic band structural features rather than background noise. This attention mechanism further enables precise localization of specific structural abnormalities within chromosomes.

Comprehensive performance evaluation of the Banding Code framework

To comprehensively evaluate the clinical effectiveness of BARCODE, we systematically validated its performance on chromosome classification and structural abnormality detection. Chromosome classification forms the foundation of karyotype analysis, because reliable chromosome identity assignment is a prerequisite for downstream

interpretation of numerical and structural abnormalities according to standardized cytogenetic nomenclature [9, 12, 13]. The model performs minimum edit distance matching between generated Banding Code sequences and standard reference sequences derived from 24 canonical ideograms. This matching process directly enables chromosome classification and structural abnormality detection. This flexible task adaptation enables a single foundational model to simultaneously support closed-set classification and open-set detection.

We selected ConvNeXtV2, FocalNet, NaViT, EfficientNetV2, ViT, FlexViT, RegNet, MaxViT, BEiT v2, and Swin Transformer as baseline models [40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50]. All baselines underwent training and evaluation on the 24-class chromosome classification task without data augmentation. Precision (Pre), Recall (Rec), and F1-score (F1) served as the core evaluation metrics. Model training utilized an in-house dataset. Figure 3a presents the comparative results. BARCODE achieves the highest scores across all three metrics. This outcome demonstrates the superior classification capability of our framework. The confusion matrix in Figure 3b further confirms that the model maintains robust classification performance for chromosome Y despite limited training samples for this class. We further conducted comparative experiments on an out-of-domain dataset derived from AutoKary2022, a densely annotated chromosome instance segmentation dataset that provides an independent benchmark for evaluating cross-domain generalization in chromosome image analysis [16]. Supplementary Figure 1a details the specific imaging parameter comparisons between datasets. The confusion matrix in Figure 3c indicates strong classification performance for chromosomes 1, 2, and 20, while chromosomes 15 and 22 exhibit relatively weaker results. Comparative analysis demonstrates that BARCODE significantly outperforms all baseline models. This outcome reveals the superior robustness of the Banding Code framework in karyotype analysis. Notably, baseline models including ViT, MaxViT, BEiT V2, and Swin Transformer show minimal classification capability on this external dataset. Chromosome categories correlate strongly with chromosome length. Distribution shifts in length between out-of-domain samples and the training set therefore introduce substantial performance deviations in generic models that rely on global pooling or fixed-size patching. Our morphology-aware full-width visual embedding module adaptively processes chromosomes of varying lengths. This design effectively bridges the distributional discrepancy and maintains consistent representation across diverse chromosome types. This representation learning strategy integrates biological priors to sustain high model performance under low-quality imaging or distribution-shifted data conditions. Such inherent robustness directly fulfills the stringent stability requirements mandated by real-world clinical workflows.

We further evaluated the zero-shot generalization capability of the model for structural abnormality detection. Structural abnormalities, including deletions, duplications, inversions, reciprocal translocations, and Robertsonian translocations, are clinically recognized cytogenetic events with important implications for prenatal diagnosis, reproductive risk assessment, and cancer cytogenetics [9, 6, 51]. Beyond gene dosage effects, such rearrangements can also perturb three-dimensional regulatory architecture by disrupting or repositioning TAD boundaries, leading to pathogenic enhancer-promoter rewiring [52, 53, 22]. Existing deep learning approaches for structural abnormality detection have generally focused on specific abnormality types or metaphase-level screening, and still struggle with rare, heterogeneous, and fine-grained band-level abnormalities [25, 18]. Furthermore, the scarcity and high heterogeneity of abnormal samples restrict the generalization of traditional supervised learning paradigms due to insufficient training data. To ensure experimental fairness, we constructed baseline models by replacing the morphology-aware full-width visual embedding in BARCODE with ViT embeddings, Swin module embeddings, or MBConv module embeddings. All configurations employed the Banding Code framework and prior-guided band enhancement strategy for data augmentation. We developed a two-tier evaluation system to comprehensively assess the clinical utility of this pipeline. The coarse-grained screening tier utilizes recall to anchor clinical sensitivity for abnormality detection. The fine-grained localization tier employs code consistency metrics to evaluate band-level detection precision. We further introduced a novel composite metric, Structured Detection Consistency Score (SDCS), to unify macroscopic detection capability and microscopic localization precision into a single quantitative measure. Figure 4a demonstrates that BARCODE achieves the highest recall rate. This performance significantly surpasses baseline pipelines utilizing generic embeddings such as ViT, Swin, or MBConv. The results confirm the superior clinical screening sensitivity of BARCODE. The MBConv-based baseline exhibits negligible detection capability. This outcome reveals a severe prediction bias toward the normal class. The model systematically misclassifies abnormal samples as normal karyotypes. This collapse stems from a failure to establish intrinsic semantic alignment between visual banding patterns and semantic Banding Code sequences. Consequently, the model loses the discriminative capacity to distinguish normal distributions from abnormal deviations. Furthermore, BARCODE delivers highly competitive performance at the fine-grained analysis level. It nevertheless accurately localizes specific abnormal bands through explicit code deviations. This mechanism guarantees the reliability of fine-grained structural detection. More critically, BARCODE achieves dominant performance on the SDCS metric. This metric jointly evaluates macroscopic screening and microscopic localization. Multi-dimensional evaluation results collectively confirm that the Banding Code-based BARCODE achieves an optimal balance between detection sensitivity and localization precision. This balance directly equips the model with exceptional zero-shot structural abnormality detection capabilities.

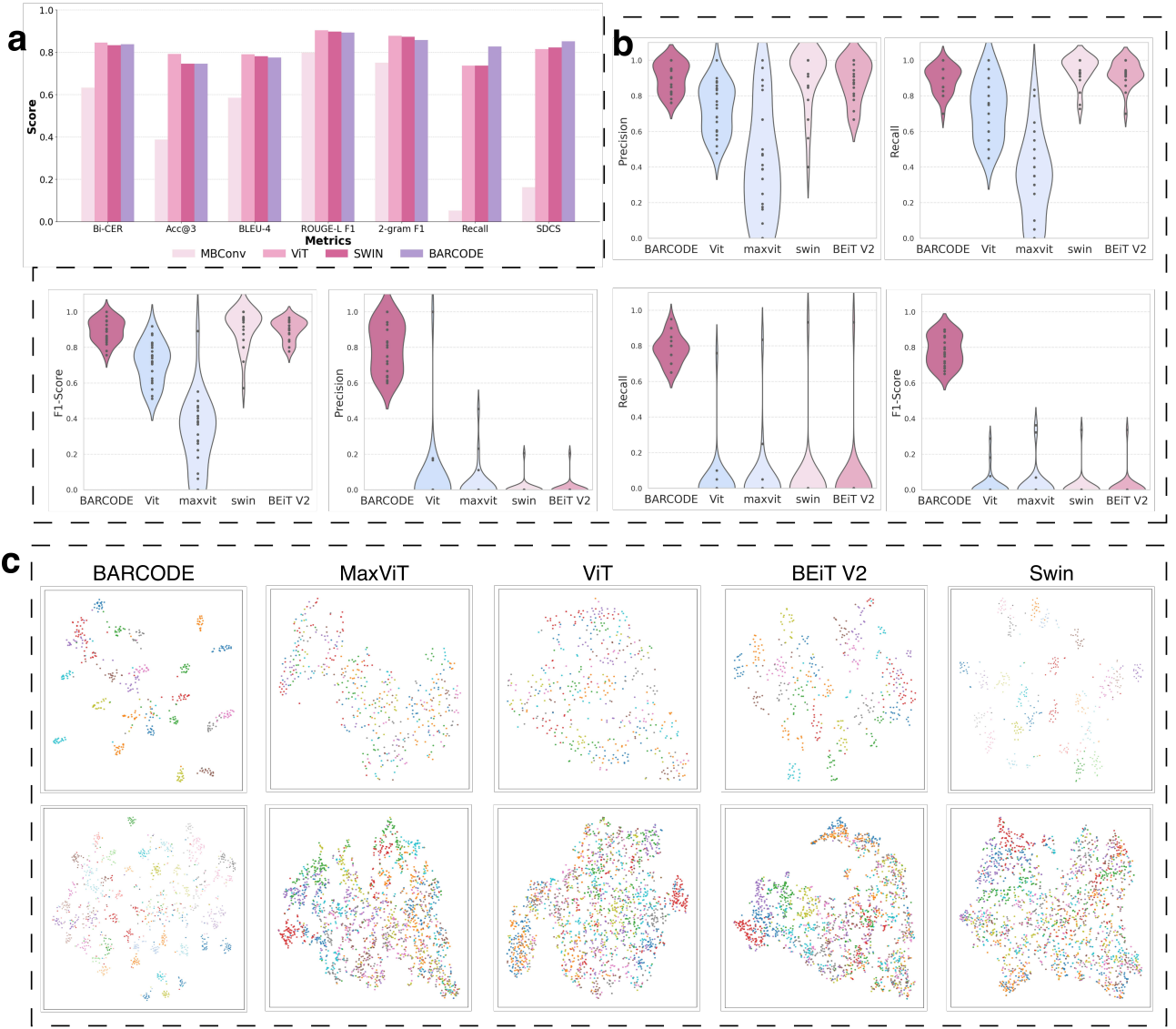


Figure 4: **Structural abnormality detection performance and feature representation analysis under the unified classification setting.** **a**, Performance comparison between BARCODE and baseline architectures on the structural abnormality detection task. **b**, Performance distributions of BARCODE and baseline models under the unified chromosome classification and structural abnormality validation task. The first three violin plots show the corresponding metrics on the internal dataset, whereas the last three violin plots show the corresponding metrics on the out-of-domain AutoKary2022 dataset. **c**, Feature embedding visualization of different models. The upper row shows results on the internal dataset, and the lower row shows results on the AutoKary2022 dataset.

Architectural Superiority Under Unified Classification and Abnormality Validation

Zero-shot detection demonstrates the generalization potential of the BARCODE. We constructed a unified classification and abnormality validation paradigm to probe the upper discriminative limit of the Banding Code representation itself. This task requires the model to simultaneously perform 24-class normal chromosome classification and structural abnormality integrity validation. We integrated the identical prior-guided band enhancement strategy across all baseline models to ensure architectural comparison fairness. This configuration directly attributes performance differences to fundamental distinctions in visual architectures and representation mechanisms rather than data distribution biases.

Figure 4b presents the experimental results. BARCODE achieves comprehensive leadership in the integrated multi-class task covering normal chromosomes and structural abnormalities. Precision, recall, and F1-score all reach optimal levels. Models such as MaxViT and ViT exhibit significantly inferior performance in this task. Out-of-domain testing further amplifies this advantage. BARCODE significantly outperforms all baseline models across every metric on the AutoKary2022 dataset. Generic visual baselines fail to demonstrate effective classification discriminability on out-of-domain data. This sharp contrast reveals that the Banding Code equips

the model with the capacity to capture essential biological features rather than merely fitting superficial texture statistics. BARCODE effectively resists out-of-domain distribution shifts and guarantees consistent performance under diverse clinical acquisition conditions. MaxViT relies on local window attention and demonstrates notably disadvantaged performance. Its overly localized feature extraction mechanism fails to capture global chromosomal banding structures even when assisted by the augmentation strategy.

We further validated representation discriminability through feature manifold visualization. Figure 4c illustrates the clustering results. BARCODE produces category feature clusters with high intra-class cohesion and clear decision boundaries. Baseline models based on MaxViT and ViT exhibit pronounced feature divergence and distribution overlap. This high-quality representation directly explains the performance advantage of BARCODE in complex classification tasks. The discrepancy becomes more prominent in the feature distribution of out-of-domain samples. Non-BARCODE visualizations lack coherent vector clustering patterns. Out-of-domain feature points deviate severely from normal cluster centers and blur classification boundaries. BARCODE leverages the Banding Code to capture essential biological features and maintains relatively compact feature cluster structures even under out-of-domain distributions. This feature-structure alignment mechanism ensures that the model executes decisions based on representations that conform to established biological priors.

Biological Rationality of the Prior-Guided Banding Augmentation Strategy

We developed a Prior-Guided Banding Augmentation Strategy (PGBA) to overcome the generalization bottleneck caused by the scarcity of structurally abnormal samples. This strategy departs from traditional geometric transformations and directly simulates clinically meaningful structural aberrations, including deletions, duplications, inversions, reciprocal translocations, and Robertsonian translocations, which are canonical classes of chromosomal rearrangements in clinical cytogenetics [9, 12, 13]. Such targeted simulation guarantees the biological plausibility of all generated samples. We performed feature manifold visualization to validate the effectiveness of this approach. The analysis directly compared simulated structural abnormalities against real pathological specimens. Figure 5a presents the t-SNE dimensionality reduction results. Simulated and real abnormal samples exhibit significant distribution overlap in the feature space. Both sample types cluster tightly within the neighborhood of their corresponding normal chromosome feature clusters. This distribution pattern conveys two critical observations. First, simulated abnormal samples successfully retain the core identity features of their source chromosomes. Second, the augmentation strategy effectively injects abnormal pattern features. These injected features align precisely with real pathological data. This feature space consistency strongly confirms the biological plausibility of the synthetic samples. The augmented data accurately reproduces the phenotypic characteristics of genuine structural abnormalities. This capability directly establishes a reliable data foundation for the zero-shot generalization of the model.

Prior-Guided Band Augmentation endows the model with zero-shot capabilities

We integrated the PGBA into the training phase to establish internal representations of normal banding grammar. This representation enables the model to identify unseen structural abnormalities as significant deviations from the normal distribution during inference. This mechanism fundamentally equips BARCODE with zero-shot detection capabilities, enabling the model to recognize diverse unknown structural abnormality types without relying on real abnormal training data. This design is particularly important in clinical cytogenetics, where balanced rearrangements, low-level mosaicism, and rare structural variants are clinically relevant but difficult to collect at scale [9, 10, 11].

Further ablation analysis quantifies the distinct contributions of each strategy component. Figure 5d illustrates these results. Translocation simulation provides the most substantial improvement in recall rates. Robertsonian translocation enhancement achieves optimal performance at the fine-grained band recognition level. This outcome indicates effective reinforcement of the model’s understanding of critical topological structures such as centromeres. Translocation simulation further delivers the highest gain across comprehensive performance metrics. This alignment reflects the high clinical prevalence of translocation abnormalities. Targeting dominant abnormal types through augmentation directly drives the expansion of the model’s generalization boundary. Conversely, baseline configurations without augmentation exhibit negligible detection capability in abnormality tasks. These baselines also demonstrate significant degradation in sequence generation metrics. Collectively, these findings confirm that the PGBA transcends conventional performance optimization techniques. The strategy operates as a core mechanism that establishes zero-shot generalization and fine-grained perceptual capabilities. This approach fundamentally resolves the generalization bottleneck imposed by the scarcity of abnormal training samples.

We evaluated the transferability of the PGBA across multiple heterogeneous visual architectures to demonstrate its broad applicability. The tested architectures encompass MaxViT, ViT, BEiT, and Swin Transformer. Figure 5b presents the empirical results. The augmentation strategy significantly improves recall rates across all tested baseline models. This enhancement directly enables zero-shot recognition of unseen abnormal types

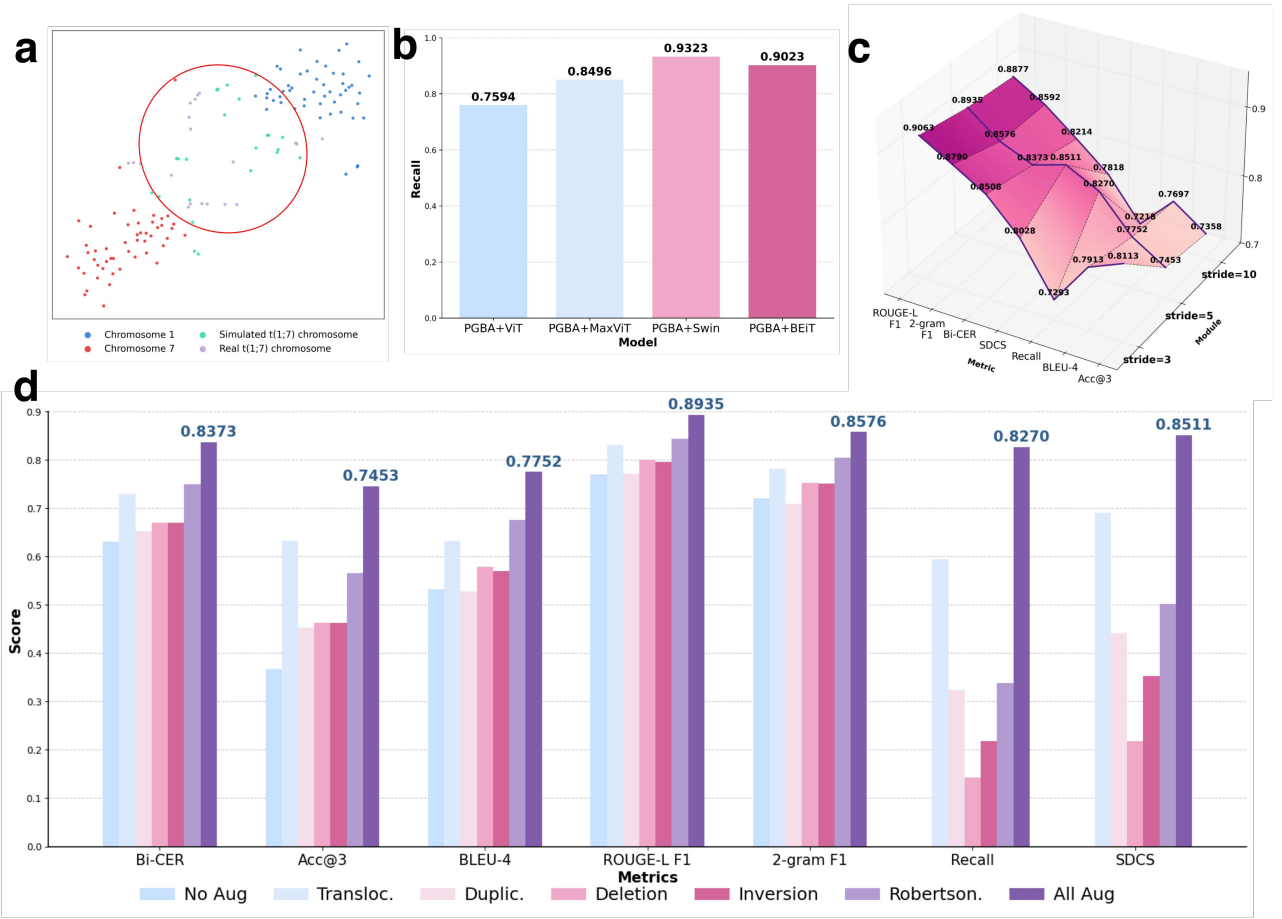


Figure 5: **Effectiveness analysis of the PGBA and the morphology-aware full-width visual embedding.** **a**, Comparison of feature manifolds between simulated and real structurally abnormal samples. **b**, Effect of the PGBA across different backbone architectures, including ViT, MaxViT, Swin and BEiT. **c**, Stride sensitivity analysis of the morphology-aware full-width visual embedding module. **d**, Ablation study of the PGBA, comparing no augmentation, individual abnormality simulation strategies, including translocation, duplication, deletion, inversion and Robertsonian translocation, and the combination of all augmentation strategy

within each architecture. The Swin Transformer framework exhibits the most pronounced performance gain. This configuration achieves optimal detection sensitivity among all evaluated models. These findings confirm the architecture-agnostic nature of the proposed strategy. The method operates independently of specific architectural designs. It fundamentally improves model generalization for abnormal distributions through the direct injection of biological prior knowledge. Consequently, the PGBA Strategy functions as a plug-and-play universal module. This component empowers diverse visual architectures to overcome clinical data scarcity. The integration of this strategy reliably enables robust zero-shot abnormal detection across heterogeneous backbones.

Morphology-Aware Full-Width Embedding significantly enhances BARCODE’s robustness

Traditional vision models, including convolutional and patch-based Transformer architectures, generally represent images through local receptive fields, fixed-size patch tokens, or global pooling operations [40, 41, 42, 46, 48, 43]. However, chromosome images exhibit a highly elongated morphology, and their clinically relevant information is organized as ordered longitudinal banding patterns. We therefore propose a morphology-aware full-width visual embedding (MAFWE) strategy to address this limitation. This approach departs from conventional methods that partition images into isolated spatial patches. It preserves the horizontal integrity of banding patterns across the entire chromosome width. Each feature unit captures the complete grayscale distribution at its corresponding vertical position. This design replicates the top-to-bottom reading logic employed by cytogeneticists during band sequence identification. The model directly converts continuous spatial image information into structurally ordered feature sequences. We conducted a hyperparameter sensitivity analysis on the longitudinal sliding stride to systematically validate the effectiveness of the morphology-aware full-width visual embedding.

Figure 5c presents the experimental results. The data reveals the specific impact of sliding stride on model

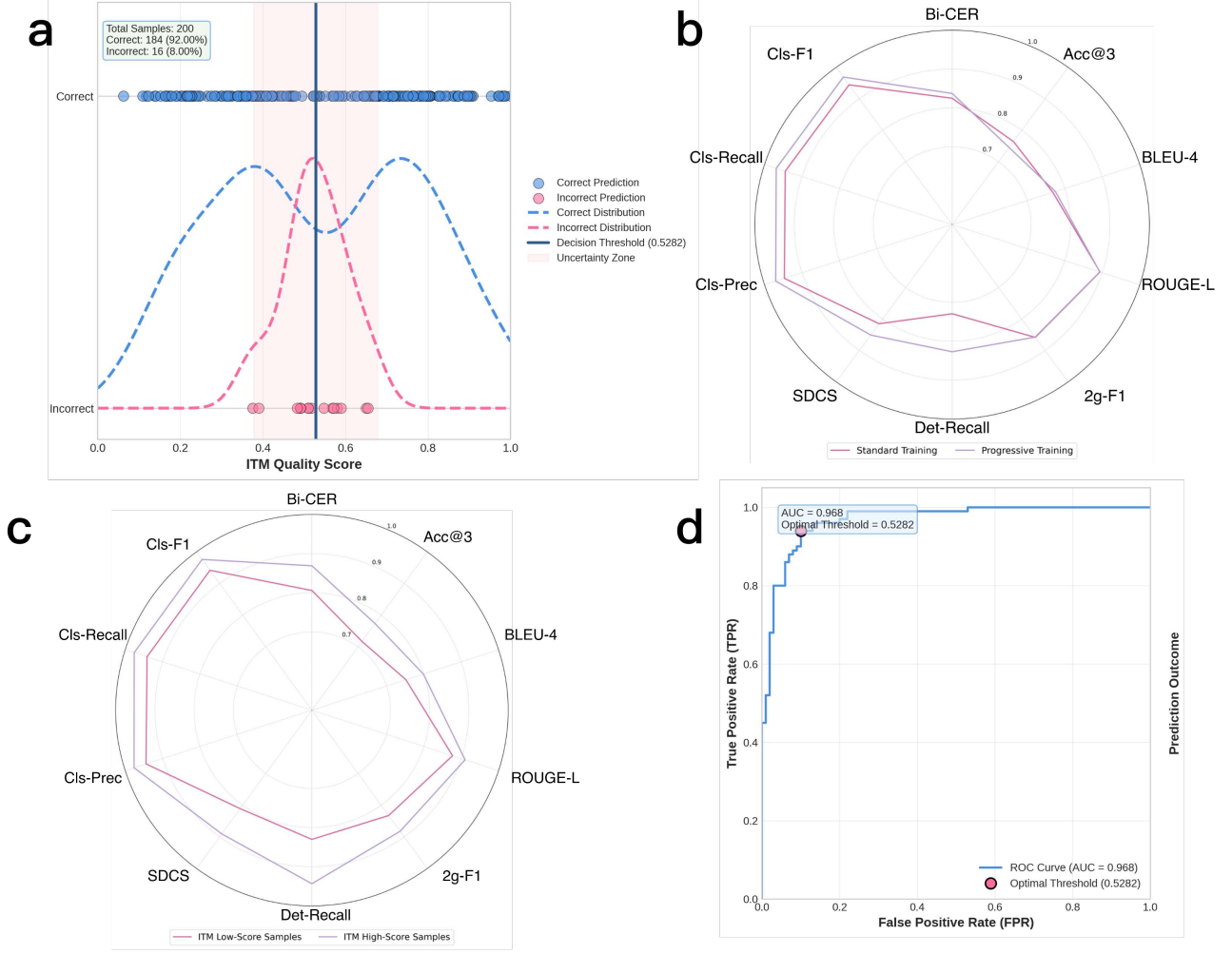


Figure 6: **ITM-based image quality stratification and uncertainty quantification.** **a**, Relationship between ITM quality scores and prediction outcomes. The score distributions of correct and incorrect predictions, the decision threshold, and the uncertainty zone are shown, indicating that low-confidence samples are mainly concentrated around the threshold region. **b**, Effect of ITM-guided progressive training on model performance. **c**, Performance comparison between ITM low-score and high-score samples. **d**, ROC analysis for ITM-based image quality stratification. The optimal threshold was determined from the ROC curve and used to distinguish high- and low-quality chromosome images.

performance. The MAFWE configuration with a stride of 3 achieves optimal performance in fine-grained band recognition tasks. Conversely, the stride-5 configuration demonstrates significantly superior recall rates for structural abnormality detection. This setting also yields the best comprehensive performance metrics. We hypothesize that the intrinsic morphological scale of chromosomal banding approximates 5 pixels under current imaging conditions. This parameter exhibits a direct physical correlation with established biological priors. Consequently, a stride of 5 aligns the sampling frequency precisely with the biological scale. This outcome further confirms the robustness and effectiveness of the MAFWE module in structural abnormality detection. Clinical diagnostics prioritize the reduction of false negative risk over fine-grained localization precision. We therefore select stride 5 as the default configuration. This choice directly maximizes clinical screening sensitivity.

Semantic Alignment-Driven Image Tiering and Robustness Enhancement

Chromosome images from clinical acquisition exhibit significant quality heterogeneity. Noise sources such as blurred banding patterns, uneven staining, or structural artifacts severely interfere with model learning of fine-grained banding features. We address this data governance challenge by leveraging the potential of the Image-Code Matching (ITM) mechanism. Similar to image-text matching and cross-modal alignment objectives in vision-language learning, the ITM module evaluates whether visual chromosome features and their corresponding Banding Code sequences are semantically consistent [37, 36, 34]. We propose an automated quality assessment strategy based on multimodal consistency scores. This strategy operates on a core hypothesis. High-quality chromosome images should exhibit strong semantic consistency with their standard Banding Code sequences. We

employ ITM similarity scores as a proxy metric for image quality. The ITM mechanism automatically partitions the dataset into high-quality and low-quality tiers. We subsequently design a progressive curriculum learning strategy that prioritizes high-quality samples after initial exposure to low-quality examples. We validated the reliability of this automated tiering strategy through consistency analysis. The analysis compared ITM-based automatic classification results against independent reviews from three senior cytogenetic experts. Figure 6a presents the agreement results. The overall concordance rate between model predictions and expert judgments reaches 92%. Discrepant samples concentrate primarily near the decision threshold. This region corresponds to the uncertainty interval in clinical interpretation. High-confidence samples exhibit a pronounced bimodal distribution. This pattern aligns precisely with expert judgment logic based on banding integrity. These results demonstrate that the ITM consistency metric accurately identifies high-quality data meeting clinical analysis standards. The evaluation logic of this metric aligns with expert criteria based on banding clarity. This ITM-based quality tiering mechanism confers explicit biological interpretability to the model. Low-quality samples typically exhibit morphological defects such as reduced banding resolution or staining artifacts. These defects decrease the semantic alignment degree between image features and Banding Code sequences. High-quality samples preserve complete banding structures. This preservation ensures cross-modal semantic consistency.

We further evaluated the impact of ITM-filtered tiered data on final model performance. Figure 6b presents the comparative results. The ITM-guided progressive curriculum learning strategy significantly optimizes model performance compared to conventional training paradigms. This strategy delivers substantial gains in recall rates and the composite SDCS metric for structural abnormality detection. All key performance metrics for chromosome classification tasks consistently surpass conventional training baselines. This performance improvement indicates that removing low-consistency samples and optimizing training order enables the model to focus on learning high-fidelity banding patterns. This focused learning directly reduces feature drift induced by noisy training data. The ITM mechanism serves dual functions within the BARCODE framework. It provides consistency constraints during inference. It simultaneously operates as a reliable data quality filter. This filtering optimizes training data distribution and fundamentally enhances the robustness and clinical applicability of the framework.

ITM-Based Uncertainty Quantification and Clinical Trust Building

The BARCODE delivers explicit code evidence as its primary output. It simultaneously achieves implicit uncertainty quantification through the ITM module. This dual functionality establishes a critical mechanism for building clinical trust. Conventional deep learning models can produce poorly calibrated or overconfident predictions, particularly under distribution shift or low-quality inputs, and therefore require explicit mechanisms for confidence estimation in safety-sensitive applications [54, 55]. The BARCODE overcomes this limitation by using the ITM similarity score as a proxy metric for prediction confidence. Figure 6c presents the correlation analysis. Model prediction accuracy demonstrates a strong positive correlation with ITM scores. Error rates for both classification and detection tasks drop significantly within the high-score subset. The low-score subset exhibits substantially higher error rates. These results confirm that the ITM metric accurately identifies model uncertainty states. This capability directly prevents overconfident predictions on high-risk samples.

Clinical workflows benefit directly from this mechanism. The model gains inherent self-awareness regarding its prediction reliability. Low ITM scores trigger a low-confidence signal when the model processes blurred banding patterns or staining artifacts. The system automatically flags these instances as cases requiring manual expert review. This routing protocol effectively prevents high-risk misdiagnoses. This transparent decision logic provides dual benefits for medical practitioners. Physicians review the final diagnostic output alongside a real-time credibility assessment. This visibility substantially reduces the trust barrier typically associated with black-box model deployment. The framework seamlessly integrates into clinical practice while maintaining diagnostic safety standards.

3 Discussion

Deep learning has advanced rapidly in medical image analysis, yet the automation of chromosome karyotype analysis remains constrained by a fundamental representation problem. Continuous visual features cannot map naturally to the discrete topological diagnostic standards defined by ISCN [12, 13]. Although previous deep learning studies have improved chromosome segmentation and classification, most approaches still operate on image-level or morphology-level representations and do not explicitly encode the biological semantics of G-banding patterns [14, 15, 16, 17]. We propose the Banding Code conceptual framework to resolve this representation dilemma. This approach discretizes continuous grayscale patterns into semantic sequences that conform to cytogenetic logic and are grounded in the biological organization of chromosome bands [29, 30, 31, 32]. The Banding Code framework fundamentally redefines the computational representation paradigm for chromosomes. Conventional strategies treat chromosomes as unordered pixel matrices and rely on passive fitting. The Banding Code explicitly maps chromosomal bands to discrete symbol units with clear biological indicators. This transformation converts

black-box visual features into structured sequences with explicit clinical semantics. This conceptual innovation drives a paradigm shift from image classification to sequence pattern recognition. BARCODE directly learns ISCN-standard banding grammar without requiring dense pixel-level annotations. Our results demonstrate that converting domain prior knowledge into computable intermediate representations provides an effective pathway. This pathway bridges the visual-semantic gap and overcomes the clinical data scarcity bottleneck. It establishes an interpretable and data-efficient foundational logic for intelligent karyotyping.

The BARCODE operates through an image-to-sequence generation paradigm. This mechanism achieves structured parsing of the intrinsic logic underlying chromosomal banding patterns. Conceptually, this design extends the image captioning paradigm from natural language generation to cytogenetic sequence generation, where the output is no longer a free-form caption but an ISCN-aligned symbolic representation of chromosome banding structure [56, 57]. Rather than passively fitting superficial textures, the model generates highly consistent codes by implicitly learning banding topological rules through Transformer-based sequence modeling, an architecture inherently adept at capturing long-range dependencies within ordered token sequences [35]. The morphology-aware full-width visual embedding module discards conventional strategies that partition images into isolated patches. It preserves the horizontal structural integrity of chromosomes. The module directly converts continuous spatial information into ordered longitudinal feature sequences. Stride sensitivity analysis identifies optimal performance at a sliding stride of 5. This sampling frequency aligns precisely with the physical scale of banding patterns under current clinical imaging conditions. The configuration enables the model to capture band arrangement rules using a top-to-bottom reading logic similar to clinical cytogeneticists. Spatial mapping of cross-attention weights further validates this mechanism. Generated abnormal Banding Code sequences concentrate precisely on actual structural breakage regions within chromosome images. Attention distributions for normal codes correspond strictly to standard band coordinates. This strategy explicitly aligns discrete banding symbols with continuous visual features. It fundamentally reconstructs the mapping relationship between algorithmic representations and clinical diagnostic logic. BARCODE consequently outputs traceable evidence chains with clear biological indicators.

BARCODE addresses abnormal sample scarcity through an integrated prior-guided banding augmentation strategy. The strategy discards arbitrary geometric transformations. It directly simulates topological structure deviations that conform to established cytogenetic rules. t-SNE visualization confirms high distribution overlap between synthesized abnormal samples and real pathological data in the feature space. The synthetic samples fully retain core identity features of their source chromosomes. The model learns only normal chromosome banding grammar during training. BARCODE identifies unseen abnormalities as significant deviations from standard distributions during inference. This mechanism directly confers robust zero-shot generalization capabilities. This capability validates the feasibility of prior-guided augmentation for rare structural variant identification.

Clinically, BARCODE serves as a complementary cytogenetic screening tool. While techniques such as CMA, CNV-seq, and QF-PCR excel at detecting submicroscopic copy-number changes, conventional karyotyping remains indispensable for assessing whole-chromosome morphology, balanced rearrangements, and low-level mosaicism [7, 10, 11]. This complementarity is particularly crucial in prenatal diagnosis and hematologic malignancy cytogenetics, where integrated testing enhances diagnostic yield and chromosome-level abnormalities retain clear clinical utility [6, 51]. By generating interpretable Banding Code sequences alongside confidence scores, BARCODE provides a scalable front-end interface for automated clinical screening and quality control. The core logic of the ITM module relies on a fundamental principle. High-quality chromosome images must maintain strong cross-modal semantic consistency with their standard Banding Code sequences. ITM similarity scores function as a proxy metric for prediction reliability. These scores achieve 92% agreement with quality assessments from senior cytogenetic experts. The confidence distribution logic aligns precisely with expert criteria based on banding integrity. The ITM mechanism drives a progressive curriculum learning strategy during training. It automatically partitions data into high- and low-quality tiers. This routing forces the model to focus on high-fidelity banding patterns and suppress noise interference. The system triggers uncertainty alerts directly upon detecting low ITM scores during inference. It automatically routes high-risk cases with blurred bands or staining artifacts to manual expert review. This quality control logic effectively prevents the overconfidence typically exhibited by black-box models under low-quality inputs. Clinicians simultaneously receive interpretable Banding Code sequences and real-time reliability assessments. BARCODE provides a secure and transparent decision interface for intelligent karyotyping. This design substantially lowers the clinical trust barriers associated with AI-assisted diagnosis.

BARCODE demonstrates significant advantages yet presents several limitations. The current model relies on Banding Code sequence deviations to indicate abnormal regions. It lacks an explicit pixel-level segmentation module. This constraint limits applicability in complex cases requiring precise breakpoint coordinate localization. In addition, the discretization granularity of the Banding Code may fail to fully characterize breakage networks during extreme topological rearrangements, especially when complex structural variants perturb regulatory architecture beyond simple band-level changes [22]. Model performance remains sensitive to sample preparation quality. ITM filtering effectively removes low-consistency data. Slight generalization deviations persist for rare samples due to skewed clinical data distributions. Future research must explore multi-scale resolution adaptation

mechanisms to mitigate these distribution biases. Subsequent work will advance along two primary pathways. The first pathway explores joint optimization of Banding Code generation and segmentation tasks. Sequence semantic constraints will assist pixel-level localization of abnormal regions. The second pathway involves constructing a cytogenetic sequence foundation model based on the Banding Code. This approach will leverage Transformer architectures for long-sequence modeling. It aims to achieve end-to-end analysis from metaphase spread images to standardized karyotype reports. We concurrently release the benchmark chromosome dataset to promote field standardization. This resource provides a reproducible evaluation baseline for future research.

In conclusion, the BARCODE establishes a credible and efficient solution to critical bottlenecks in clinical karyotyping. It achieves this goal through high-granularity, band-level chromosome analysis based on the Banding Code representation.

4 Methods

Banding Code Design

The Banding Code is derived from the ISCN standard G-banding ideograms of human chromosomes, as shown in Supplementary Figure 1b, which represent an established expert consensus for discretizing continuous chromosomal grayscale patterns into biologically meaningful band categories [3, 12, 13]. Building on this foundation, we formalize this discretization as a symbolic sequence representation to enable band-level fine-grained analysis. Specifically, each position along the chromosome is assigned one of four tokens: 1 denotes G-positive dark bands, which are generally AT-rich, gene-poor, and late-replicating; 2 denotes transitional gray bands; 3 denotes G-negative light bands, which are generally GC-rich, gene-rich, and early-replicating; and 4 marks the centromere position [29, 30, 31, 32]. The initial token assignments are derived from ISCN ideograms and subsequently refined through manual boundary correction by experienced cytogeneticists to account for imaging-specific variations. The resulting Banding Code sequences for all 24 chromosome classes are provided in Supplementary Figure 21. In BARCODE, these sequences serve as the generation targets of the Image-to-Sequence task, providing an interpretable intermediate representation that bridges low-level visual features and high-level diagnostic standards.

Morphology-Aware Full-Width Visual Embedding

To address the structural misalignment between morphological features and visual representations in conventional models, we propose a Morphology-Aware Full-Width Visual Embedding (MAFWE) module. The core of this module is a full-width convolutional operation in which the kernel dimensions are set to $k \times W$, with the width dimension strictly aligned to the input image width W . Convolution is performed exclusively along the longitudinal axis of the chromosome with a vertical stride s , while the horizontal stride is set equal to W . This configuration ensures that the output width is always 1, effectively reducing the two-dimensional convolution to a one-dimensional sliding operation along the chromosomal long axis. Given an input chromosome grayscale image $\mathbf{X} \in \mathbb{R}^{H \times W}$, the i -th feature token $\mathbf{t}_i \in \mathbb{R}^d$ is defined as:

$$\mathbf{t}_i = \text{Norm}(\mathbf{W}_c * \mathbf{X}_{[i:s : i:s+k, 0:W]}), \quad i = 0, 1, \dots, N-1 \quad (1)$$

where $\mathbf{W}_c$ denotes the learnable convolutional parameters, $\text{Norm}(\cdot)$ denotes layer normalization, and the sequence length is given by $N = \lfloor (H - k)/s \rfloor + 1$. The resulting token sequence $\{\mathbf{t}_0, \mathbf{t}_1, \dots, \mathbf{t}_{N-1}\}$ is augmented with sinusoidal positional encodings and subsequently fed into the Transformer encoder.

Unlike conventional patch-based embedding methods that partition images into isolated spatial regions, this design integrates the complete cross-sectional grayscale distribution at each sliding step, ensuring that every feature token captures a holistic representation of its corresponding longitudinal position. This prevents the fragmentation of banding features caused by arbitrary spatial partitioning. Furthermore, the resulting token sequence follows the natural longitudinal ordering of the chromosome, closely mirroring the top-to-bottom band identification logic employed by cytogeneticists, thereby providing biologically grounded visual anchors for the subsequent autoregressive generation of Banding Code sequences.

Prior-Guided Banding Augmentation Strategy

To overcome the generalization bottleneck caused by the scarcity of structurally abnormal samples, we developed a Prior-Guided Banding Augmentation Strategy. Unlike conventional geometric transformations, this strategy aims to simulate clinically meaningful structural aberrations, thereby ensuring the biological plausibility of all synthesized samples. Specifically, we first localize discrete banding regions within chromosome images through conventional grayscale profile analysis combined with ISCN banding distribution priors. Structured operations are then applied to the extracted banding units: inter-chromosomal banding exchange is used to simulate translocations, and additionally, for acrocentric chromosomes prone to Robertsonian translocations (chromosomes 13,

14, 15, 21, and 22), the localized centromeres are fused to synthesize Robertsonian translocation samples. Intra-chromosomal self-exchange is applied to simulate inversions, while the duplication and deletion of banding regions correspond to duplications and deletions, respectively. This banding semantics-based augmentation approach ensures that synthesized abnormal samples are highly consistent with genuine pathological specimens in terms of both topological structure and textural characteristics. By introducing such biologically grounded distributional deviations during the learning phase, the model establishes an internal representation of normal banding grammar, enabling it to recognize unseen structural abnormalities as significant deviations from the normal distribution during inference.

Image-Text Matching and Data Quality Stratification Strategy

The image-text matching model consists of two submodules: an Image-grounded Text Encoder and an Image-grounded Text Decoder, both of which share the parameters of all Cross-attention layers and feed-forward networks (FFN). This parameter-sharing mechanism compels the encoder and decoder to operate within a unified visual-semantic space, ensuring that the ITM discrimination objective and the sequence generation objective share a consistent image feature representation, while simultaneously reducing the risk of overfitting[?, 36]. The Image-grounded Text Encoder models the Banding Code text sequence through bidirectional self-attention layers to extract its global semantic representation, and adopts the ITM loss as its optimization objective, formulating the image-code matching relationship as a binary classification task. A hard negative mining strategy is employed during training to obtain more informative negative samples, thereby enhancing the discriminative power of the image-text alignment representation. The Image-grounded Text Decoder generates Banding Code sequences conditioned on chromosome image features, modeling previously generated tokens via a causal self-attention mechanism and autoregressively predicting subsequent Banding Codes, with a token-level cross-entropy loss as the optimization objective. This enables the model to learn to generate Banding Code sequences conforming to clinical banding distributions given the chromosome image and the historical token prefix. It is worth noting that the ITM model is pre-trained on an independent dataset subjected to additional data augmentation, completely isolated from the formal training data, thereby preventing data contamination during quality assessment.

The ITM scores produced during the above process simultaneously serve as a proxy metric for image quality. We hypothesize that high-quality chromosome images should exhibit strong semantic consistency with their corresponding standard Banding Code sequences, whereas quality defects such as blurred banding patterns, uneven staining, or structural artifacts reduce the degree of image-text semantic alignment, thereby yielding lower ITM scores. To determine the optimal quality stratification threshold, we randomly sampled a representative subset from the formal training set and recruited experienced cytogeneticists to perform binary quality annotations. Following normalization of ITM scores across all samples, an ROC curve was constructed based on this annotated subset, and the optimal classification threshold was determined using the Youden index. This threshold was subsequently applied to the entire training dataset to automatically stratify samples into high-quality and low-quality tiers. Building on this, we designed a progressive curriculum learning strategy that utilizes only the Encoder and the Image-grounded Text Decoder. In the first stage, low-quality samples are introduced to enhance the model’s robustness to noise. In the second stage, the Encoder is frozen, and fine-grained training is conducted using high-quality samples. This approach guides the model to progressively focus on learning high-fidelity banding features, effectively suppressing feature drift induced by noisy data.

Information of Metrics

For general sequence generation quality, we report BLEU-4, ROUGE-L F1 and 2-gram F1, which are widely used metrics for evaluating n-gram overlap and longest-common-subsequence similarity in generated sequences [38, 39]. In the meanwhile, we introduce three task-specific metrics for structural abnormality detection: the Bidirectional Coding Correctness Rate (Bi-CER), the Fault-Tolerant Accuracy (Acc@3), and the Structured Detection Consistency Score (SDCS).

Bi-CER and Acc@3 are both grounded in the Banding Code matching algorithm. Let the predicted Banding Code be $Q = q_1 q_2 \dots q_m$, and let the set of 24 standard normal Banding Codes form the candidate set:

$$\mathcal{C} = \{C^{(1)}, \dots, C^{(24)}\}, \quad C^{(j)} = c_1^{(j)} \dots c_{n_j}^{(j)} \quad (2)$$

For each candidate $C^{(j)}$, we construct a cost matrix $D^{(j)}(i, l)$ representing the minimum edit cost to transform the first i characters of Q into the first l characters of $C^{(j)}$. The initialization conditions are:

$$D^{(j)}(0, 0) = 0, \quad D^{(j)}(i, 0) = i, \quad D^{(j)}(0, l) = l \quad (3)$$

The recurrence relation is:

$$D^{(j)}(i, l) = \min \begin{cases} D^{(j)}(i-1, l-1) + \delta(q_i, c_l^{(j)}) \\ D^{(j)}(i-1, l) + 1 \\ D^{(j)}(i, l-1) + 1 \end{cases} \quad (4)$$

where the match cost function δ is defined as:

$$\delta(q_i, c_l^{(j)}) = \begin{cases} 0, & \text{if } q_i = c_l^{(j)} \\ 1, & \text{if } q_i \neq c_l^{(j)} \end{cases} \quad (5)$$

The total matching cost for candidate $C^{(j)}$ is $\text{Cost}(Q, C^{(j)}) = D^{(j)}(m, n_j)$, and the optimal matched class is:

$$C^{(*)} = \arg \min_{j=1}^{24} \text{Cost}(Q, C^{(j)}) \quad (6)$$

If $\text{Cost}(Q, C^{(*)}) < \tau$, the sample is classified as normal; otherwise, a structural abnormality is detected.

Since chromosome images may be captured in either orientation, the predicted Banding Code sequence may appear in forward or reverse order. We therefore compute the edit distance in both directions and take the minimum as the final bidirectional edit distance:

$$d_{\text{bi}}(Q, G) = \min(\text{Cost}(Q, G), \text{Cost}(Q, G^{\text{rev}})) \quad (7)$$

where G^{rev} denotes the reverse of the ground-truth banding code G . Building on this, Bi-CER is defined as the normalized bidirectional code correctness rate:

$$\text{Bi-CER} = 1 - \frac{d_{\text{bi}}(Q, G)}{\max(|Q|, |G|)} \quad (8)$$

A higher Bi-CER indicates greater band-level agreement between the predicted and ground-truth sequences. Acc@3 further introduces a fault-tolerant mechanism, defined as the proportion of samples whose bidirectional edit distance does not exceed 3:

$$\text{Acc@3} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}[d_{\text{bi}}(Q_i, G_i) \leq 3] \quad (9)$$

This metric characterizes model robustness under fault-tolerant conditions, consistent with the clinical reality that minor band boundary deviations do not compromise overall diagnostic conclusions.

SDCS (Structured Detection Consistency Score) is a composite metric designed to jointly evaluate macroscopic screening sensitivity and microscopic localization precision. We first aggregate the five fine-grained sequence metrics — Bi-CER, Acc@3, BLEU-4, ROUGE-L F1, and 2-gram F1 — into a unified fine-grained score SeqScore via harmonic mean:

$$\text{SeqScore} = \frac{5}{\frac{1}{\text{Bi-CER}} + \frac{1}{\text{Acc@3}} + \frac{1}{\text{BLEU-4}} + \frac{1}{\text{ROUGE-L}} + \frac{1}{\text{2-gram F1}}} \quad (10)$$

The harmonic mean is adopted over the arithmetic mean to penalize imbalance across fine-grained metrics, ensuring consistently high performance across all sequence quality dimensions. The final SDCS is then computed as the harmonic mean of the coarse-grained detection Recall and SeqScore:

$$\text{SDCS} = \frac{2 \times \text{Recall} \times \text{SeqScore}}{\text{Recall} + \text{SeqScore}} \quad (11)$$

This hierarchical harmonic mean design assigns equal 50% weight to screening sensitivity and band-level localization precision, preventing strong performance on one dimension from masking deficiencies on the other, and thereby providing a reliable quantitative benchmark for clinically-oriented comprehensive evaluation.

Information of Baselines

To comprehensively evaluate the clinical effectiveness of BARCODE, we systematically selected and constructed various baseline models across different tasks. For the task of chromosome classification, we selected diverse mainstream vision models as baselines, including ConvNeXtV2, FocalNet, NaViT, Efficient Net V2, ViT, FlexViT, RegNet, MaxViT, BeiTv2, and Swin Transformer[40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50]. Notably, all these generic visual baseline models are sourced from the TIMM (PyTorch Image Models) library. These models primarily rely on global pooling or fixed-size patching mechanisms for feature representation. For the task of

structural abnormality detection, to ensure experimental fairness, we constructed baseline models by replacing the morphology-aware full-width visual embedding in BARCODE with generic embeddings. Specifically, we utilized ViT, Swin, or MBConv module embeddings as alternative visual encoders. All baseline configurations retained the Banding Code framework and employed the same prior-guided band enhancement strategy for data augmentation. For the task of unified classification and abnormality validation, to probe the discriminative limits of different architectures, we integrated the identical prior-guided band enhancement strategy across all TIMM-based baseline models, including MaxViT, ViT, BEiT, and Swin Transformer. This setup ensures architectural comparison fairness, directly attributing performance differences to fundamental distinctions in visual architectures and representation mechanisms rather than data distribution biases.

Dataset and annotation

Our data were all obtained from the International Peace Maternity and Child Health Hospital of China Welfare Institute, and the public datasets AutoKary2022[16]. The hospital dataset, collected from multiple centers, comprises 83,812 normal chromosome images and 133 structurally abnormal chromosome images. AutoKary2022, originally intended for segmentation tasks, contains lower-quality images; 2,400 segmented images were randomly selected as an out-of-domain test set. During training, only normal chromosomes were used and split into training and validation subsets at a 9:1 ratio. For structural abnormality detection, an independent set of real abnormal chromosomes was used to evaluate the model’s zero-shot capability. Additionally, to assess consistency across banding resolutions, a dataset comprising 24 classes of normal chromosomes at 300-, 400-, and 550-band levels was constructed.

Input images were converted to grayscale and white borders were removed, and each chromosome instance was resized to a fixed width of 40 pixels while preserving aspect ratio. During training, random horizontal and vertical flipping, brightness perturbation, and contrast adjustment were applied, with vertical flipping also reversing the Banding Code sequence to maintain orientation consistency. Visual embeddings were extracted using the morphology-aware full-width embedding module with a kernel size of $\text{stride} \times 40$ and a longitudinal stride of 5. The resulting token sequence was fed into a Transformer encoder-decoder (hidden dimension 256, 12 encoder and decoder layers, 8 attention heads, feed-forward dimension 2048, dropout 0.1, pre-normalization). Decoder outputs were mapped to a Banding Code vocabulary of 7 tokens (including padding, Banding Code tokens, start and end tokens) via a three-layer MLP, with a maximum sequence length of 32. The model was optimized using AdamW (base learning rate 1×10^{-4} , weight decay 1×10^{-4} ; bias and normalization parameters excluded from decay). Training followed a two-stage scheme: Stage 1 for 20 epochs with cross-entropy loss and label smoothing 0.05; Stage 2 for 30 epochs with learning rate reduced to 1×10^{-5} , applying module-specific learning-rate multipliers and freezing the patch embedding, Transformer encoder, and embedding layers for initial epochs before joint fine-tuning. A StepLR scheduler and gradient clipping (max norm 0.1) were used. Batch size was 64 with 32 data-loading workers, trained on an RTX 4090. Random seed was fixed at 42. Validation loss was monitored each stage, and the lowest-loss checkpoint was saved. The final model was selected based on structural abnormality detection performance.

Data preprocessing

During data preprocessing, individual chromosome instances were first extracted from metaphase images. Specifically, Mask R-CNN was used to perform preliminary instance segmentation on raw metaphase spreads, generating initial chromosome masks and class predictions. The resulting segmentation boundaries and chromosome labels were subsequently reviewed and corrected by experienced cytogeneticists to ensure accurate instance extraction and consistent chromosome annotation. For each extracted chromosome instance, geometric straightening was performed using thin-plate spline (TPS) transformation to reduce morphological variability caused by bending, curvature and local deformation. The detailed workflow is illustrated in Supplementary Figure 1c. First, the Zhang–Suen thinning algorithm was applied to the chromosome mask to extract its skeletal structure. Because skeletons derived from clinical chromosome images may contain spurious branches caused by local adhesion, irregular boundaries or image noise, we further analysed the geometric relationships between skeleton points and chromosome contours to identify and remove redundant branches. Local bifurcations were then merged to obtain a continuous, branch-free primary skeleton representing the longitudinal axis of the chromosome. Next, normal directions were computed at each point along the primary skeleton, and their intersections with the chromosome boundaries were identified to define cross-sectional correspondences along the chromosome body. Based on the correspondence between boundary sampling points in the original curved configuration and their target positions in the straightened configuration, pairs of TPS control points were constructed and used to estimate a non-rigid geometric transformation. This transformation was then applied to the original chromosome instance image, yielding a straightened and spatially standardized chromosome representation for subsequent Banding Code generation and band-level analysis.

5 Data Availability

Our dataset will be made publicly available at <https://drive.google.com/drive/folders/160S-Vjax308jdvvh2Vahu2qM95WctRIK> and the AutoKary2022 dataset is available at <https://github.com/wangjuncongyu/chromosome-instance-segmentation-dataset>.

6 Code Availability

The codes of BARCODE can be found in <https://github.com/qiuyuzhe1111/BARCODE-A-Chromosome-Karyotype-Analysis-Model-Based-on-a-Banding-Code-Framework/tree/mains>.

7 Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 62227811 and 62473255), Special Research Program on Smart Healthcare of Shanghai Municipal Health Commission (Grant No. 2025ZHYL036), and the Program Medical-Engineering Interdisciplinary Research Fund of Shanghai Jiao Tong University (Grant No. YG2025LC11). We thank the Department of Gynecology of the International Peace Maternity and Child Health Hospital for providing data support. This study was approved by the Ethics Committee of the International Peace Maternity and Child Health Hospital of Shanghai Jiao Tong University School of Medicine (GKLW-A-2025-109-01). All participants signed the informed consent.

8 Author contributions

ZH Qiu conceived the study, designed the methodology, performed data preprocessing, constructed the model, conducted the experiments, and wrote the original draft. WJ Hu conceived the study, performed patient screening and enrollment, collected and annotated the clinical data, validated the data, and wrote the original draft. H Cheng supervised the computational workflow and provided methodological guidance. JW Cao contributed to study conception and experimental execution. CC Gong performed data preprocessing and data organization. SY Li performed patient screening and enrollment, collected and annotated the clinical data, and validated the data. HD Pan supervised the computational workflow and provided methodological guidance. CD Li performed data preprocessing and data organization. XP Guan revised the manuscript. H Xu supervised the computational workflow, provided methodological guidance, validated the data, revised the manuscript, and approved the final version. KJ Wu supervised the computational workflow, provided methodological guidance, revised the manuscript, and approved the final version. All authors reviewed and approved the final manuscript.

9 Competing interests

The authors declare no competing interests.

References

- [1] Aynaci, S. *et al.* Chromosomal abnormalities in couples with recurrent pregnancy loss: a 16-year cross-sectional study of 4030 cases from turkey. *Annals of Saudi Medicine* **45**, 154–164 (2025).
- [2] Kunda, S. *et al.* Spectrum and clinical reproductive significance of cytogenetic abnormalities in infertility and recurrent early pregnancy loss: A five-year retrospective study of 10,285 cases. *Cureus* **18** (2026).
- [3] Seabright, M. A rapid banding technique for human chromosomes. *The Lancet* **2**, 971–972 (1971).
- [4] Sumner, A. T., Evans, H. J. & Buckland, R. A. New technique for distinguishing between human chromosomes. *Nature New Biology* **232**, 31–32 (1971).
- [5] Steele, M. W. & Breg, W. R. Chromosome analysis of human amniotic-fluid cells. *The Lancet* **1**, 383–385 (1966).
- [6] Rowley, J. D. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290–293 (1973).
- [7] Wapner, R. J. *et al.* Chromosomal microarray versus karyotyping for prenatal diagnosis. *New England Journal of Medicine* **367**, 2175–2184 (2012).

- [8] Levy, B. & Wapner, R. Prenatal diagnosis by chromosomal microarray analysis. *Fertility and Sterility* **109**, 201–212 (2018).
- [9] Gardner, R. J. M., Sutherland, G. R. & Shaffer, L. G. *Chromosome Abnormalities and Genetic Counseling* (Oxford University Press, Oxford, 2012), 4 edn.
- [10] Liu, J.-p., Wang, S.-B., Luo, L. & Guo, Y.-m. Improving prenatal diagnosis with combined karyotyping, CNV-seq and QF-PCR: a comprehensive analysis of chromosomal abnormalities in high-risk pregnancies. *Frontiers in Genetics* **15**, 1517270 (2025).
- [11] Xue, S. *et al.* Clinical application of chromosome microarray analysis and karyotyping in prenatal diagnosis in northwest china. *Frontiers in Genetics* **15**, 1347942 (2024).
- [12] McGowan-Jordan, J., Hastings, R. J. & Moore, S. (eds.) *ISCN 2020: An International System for Human Cytogenomic Nomenclature* (S. Karger AG, Basel, 2020).
- [13] Hastings, R. J., Moore, S. & Chia, N. (eds.) *ISCN 2024: An International System for Human Cytogenomic Nomenclature* (S. Karger AG, Basel, 2024).
- [14] Lin, C. *et al.* CIR-Net: automatic classification of human chromosome based on Inception-ResNet architecture. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **19**, 1285–1293 (2022).
- [15] Huang, R. *et al.* A clinical dataset and various baselines for chromosome instance segmentation. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **19**, 31–39 (2022).
- [16] You, D. *et al.* AutoKary2022: A large-scale densely annotated dataset for chromosome instance segmentation. In *2023 IEEE International Conference on Multimedia and Expo (ICME)* (2023).
- [17] Tseng, J.-J. *et al.* An open dataset of annotated metaphase cell images for chromosome identification. *Scientific Data* **10**, 104 (2023).
- [18] Kang, S. *et al.* Chromosome analysis method based on deep learning: counting chromosomes and detecting abnormal chromosomes. *Biomedical Signal Processing and Control* **88**, 105891 (2024).
- [19] Nagaoka, S. I., Hassold, T. J. & Hunt, P. A. Human aneuploidy: mechanisms and new insights into an age-old problem. *Nature Reviews Genetics* **13**, 493–504 (2012).
- [20] McCoy, R. C. *et al.* Meiotic and mitotic aneuploidies drive arrest of in vitro fertilized human preimplantation embryos. *Genome medicine* **15**, 77 (2023).
- [21] Pylyp, L. Y. *et al.* Chromosomal abnormalities in products of conception of first-trimester miscarriages detected by conventional cytogenetic analysis: a review of 1000 cases. *Journal of assisted reproduction and genetics* **35**, 265–271 (2018).
- [22] Spielmann, M., Lupiáñez, D. G. & Mundlos, S. Structural variation in the 3d genome. *Nature Reviews Genetics* **19**, 453–467 (2018).
- [23] Hu, X. *et al.* Classification of metaphase chromosomes using deep convolutional neural network. *Journal of Computational Biology* **26** (2019).
- [24] Qin, Y. *et al.* Varifocal-Net: a chromosome classification approach using deep convolutional networks. *IEEE Transactions on Medical Imaging* **38**, 2569–2581 (2019).
- [25] Yan, J., Tucci, E. & Jaffe, N. Detection of t(9;22) chromosome translocation using deep residual neural network. *Journal of Computer and Communications* **7**, 102–111 (2019).
- [26] He, W. *et al.* Feature fusion classifier with dynamic weights for abnormality detection of amniotic fluid cell chromosome. *IEEE Access* **11**, 31755–31766 (2023).
- [27] Chen, S., Hu, T., Li, N., Gao, X. & Yu, Y. Br-chromnet: Banding resolution localization of chromosome structural abnormality with conditional random field. *Journal of molecular biology* **436**, 168733 (2024).
- [28] Uzolas, L., Rico, J., Coupé, P., SanMiguel, J. C. & Cserey, G. Deep anomaly generation: An image translation approach of synthesizing abnormal banded chromosome images. *IEEE Access* **10**, 59090–59098 (2022).
- [29] Holmquist, G., Gray, M., Porter, T. & Jordan, J. Characterization of Giemsa dark- and light-band DNA. *Cell* **31**, 121–129 (1982).

- [30] Craig, J. M. & Bickmore, W. A. Chromosome bands—flavours to savour. *BioEssays* **15**, 349–354 (1993).
- [31] Niimura, Y. & Gojobori, T. In silico chromosome staining: reconstruction of Giemsa bands from the whole human genome sequence. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 797–802 (2002).
- [32] Gilbert, N. *et al.* Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* **118**, 555–566 (2004).
- [33] Wang, J. *et al.* Git: A generative image-to-text transformer for vision and language. *arXiv preprint arXiv:2205.14100* (2022).
- [34] Li, J., Li, D., Xiong, C. & Hoi, S. BLIP: bootstrapping language-image pre-training for unified vision-language understanding and generation. In *Proceedings of the 39th International Conference on Machine Learning*, 12888–12900 (2022).
- [35] Vaswani, A. *et al.* Attention is all you need. In *Advances in Neural Information Processing Systems*, vol. 30 (2017).
- [36] Li, J. *et al.* Align before fuse: vision and language representation learning with momentum distillation. In *Advances in Neural Information Processing Systems*, vol. 34, 9694–9705 (2021).
- [37] Radford, A. *et al.* Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, 8748–8763 (2021).
- [38] Papineni, K., Roukos, S., Ward, T. & Zhu, W.-J. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311–318 (2002).
- [39] Lin, C.-Y. ROUGE: a package for automatic evaluation of summaries. In *Text Summarization Branches Out*, 74–81 (2004).
- [40] Dosovitskiy, A. *et al.* An image is worth 16x16 words: transformers for image recognition at scale. In *International Conference on Learning Representations* (2021).
- [41] Liu, Z. *et al.* Swin transformer: hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012–10022 (2021).
- [42] Liu, Z. *et al.* A ConvNet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11976–11986 (2022).
- [43] Woo, S. *et al.* ConvNeXt V2: co-designing and scaling ConvNets with masked autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16133–16142 (2023).
- [44] Tan, M. & Le, Q. V. EfficientNetV2: smaller models and faster training. In *Proceedings of the 38th International Conference on Machine Learning*, 10096–10106 (2021).
- [45] Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K. & Dollár, P. Designing network design spaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10428–10436 (2020).
- [46] Tu, Z. *et al.* MaxViT: multi-axis vision transformer. In *European Conference on Computer Vision*, 459–479 (Springer, 2022).
- [47] Yang, J., Li, C., Dai, X. & Gao, J. Focal modulation networks. In *Advances in Neural Information Processing Systems*, vol. 35, 4203–4217 (2022).
- [48] Peng, Z., Dong, L., Bao, H., Ye, Q. & Wei, F. BEiT v2: masked image modeling with vector-quantized visual tokenizers. *arXiv preprint arXiv:2208.06366* (2022). 2208.06366.
- [49] Beyer, L. *et al.* FlexiViT: one model for all patch sizes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14496–14506 (2023).
- [50] Dehghani, M. *et al.* Patch n’ pack: NaViT, a vision transformer for any aspect ratio and resolution. In *Advances in Neural Information Processing Systems*, vol. 36 (2023).
- [51] Kou, F., Wu, L., Ren, X. & Yang, L. Chromosome abnormalities: new insights into their clinical significance in cancer. *Molecular Therapy: Oncolytics* **17**, 562–570 (2020).

- [52] Lupiáñez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
- [53] Lupiáñez, D. G., Spielmann, M. & Mundlos, S. Breaking TADs: how alterations of chromatin domains result in disease. *Trends in Genetics* **32**, 225–237 (2016).
- [54] Guo, C., Pleiss, G., Sun, Y. & Weinberger, K. Q. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, 1321–1330 (2017).
- [55] Kendall, A. & Gal, Y. What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems*, vol. 30 (2017).
- [56] Vinyals, O., Toshev, A., Bengio, S. & Erhan, D. Show and tell: a neural image caption generator. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3156–3164 (2015).
- [57] Xu, K. *et al.* Show, attend and tell: neural image caption generation with visual attention. In *Proceedings of the 32nd International Conference on Machine Learning*, 2048–2057 (2015).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Figuredata.xlsx](#)
- [nccode.zip](#)
- [NCSupplementary.pdf](#)